





الجمهورية العربية السورية

وزارة التعليم العالي والبحث العلمي

جــامعـة تشــريــن

محاضرات

في

مناهج وأساليب البحث العلمي

الدكتور	الدكتورة
إبراهيم محمد العلي	**يسيرة دريباتي**

أستاذان في قسم الإحصاء والبرمجة

كلية الاقتصاد- جامعة تشرين- سورية

اللاذقية – سورية

2021 م – 1442 هـ

الفهرس العام

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

الموضوع	الصفحة

المقدمة :

لم يعد البحث العلمي عملية كيفية، يقوم بها الباحث بشكل غير منظم، بل أضحى وسيلة أساسية لاكتشاف الحقائق وإنجاز الإختراعات المختلفة بطرائق منهجية وأساليب إحصائية. لذلك قمنا بكتابة هذه المحاضرات المختصة لتكون دليلاً مختصراً لطلاب الدراسات العليا في كلية الاقتصاد وغيرها حول مناهج وأساليب البحث العلمي، ولقد جمعنا موضوعاتها ضمن جزأين كما يلي:

الجزء الأول: ويتناول القضايا النظرية والمنهجية للبحث العلمي ويتضمن الفصول التالية:

- الفصل الأول: ويتضمن الأسس النظرية للبحث العلمي وشرحاً لمفهومه وخصائصه وميادين تطبيقه.
- الفصل الثاني: ويضم أهم المناهج المستخدمة في البحث العلمي، وفيه نتعرض إلى المناهج الأساسية التالية: الاستقرائي والاستنتاجي والتاريخي والتجريبي والوصفي والمسحي وتحليل المضمون ودراسة الحالة وتحليل النظم.
- الفصل الثالث: ويحتوي على عرض لخطوات مرحلة التحضير لإعداد البحث العلمي، وتبدأ باختيار الموضوع وتحديد المشكلة وأهميتها وأهداف البحث ومجاله ومجتمعه وعينته ومتحولاته وصياغة فروضه ثم تحديد مناهجه وأدواته ووضع مخططه ومراجعه ...الخ .
- الفصل الرابع: ويتضمن عرضاً لخطوات مرحلة جمع البيانات وشرحاً لطرائق جمعها كالقياس والاستخلاص والملاحظة والمقابلة والاستبيان .
- الفصل الخامس: ويتعرض للأسس العامة لكتابة البحث أو الرسالة أو الأطروحة. ويلخص قواعد وشروط كتابة التقرير النهائي للبحث، وهي القواعد والشروط المتفق عليها في جميع مراكز البحوث العلمية .

الجزء الثاني: ويتناول أهم الأساليب الإحصائية المستخدمة في البحث العلمي ويتضمن الفصول التالية :

- الفصل الأول: وفيه نستعرض أهم أدوات معالجة البيانات كالترتيب والتبويب ووسائل عرض البيانات.
- الفصل الثاني: وفيه نتعرض لحساب مقاييس النزعة المركزية والتشتت كالمتوسط الحسابي والوسيط والمنوال والتباين والانحراف المعياري ...الخ.
- الفصل الثالث: ويتناول التوزيعات الاحتمالية الأكثر استخداماً في البحوث العلمية كالتوزيع الطبيعي العام والمعياري وتوزيع ستيودنت وتوزيع كاي مربع وتوزيع فيشر.
- الفصل الرابع: ويشمل عرضاً مختصراً للعينات العشوائية وطرائق التقدير الاحصائي لكل من المتوسط والانحراف المعياري والخطأ المعياري وغيرها .
- الفصل الخامس: ويتناول كيفية تصميم الاستبيان واختبار مصداقيته وحساب معاملات ثبات إجاباته.

- الفصل السادس: ويعالج قضايا اختبار الفرضيات حول متوسط المجتمع والنسبة فيه وحول الفرق بين متوسطي مجتمعين ونسبتين فيهما ....الخ.
- لفصل السابع: ويتناول قضايا تحليل التباين البسيط (باتجاه واحد وباتجاهين) والأساليب المتعلقة به (كالمربع اللاتيني وتحليل التغاير).
- الفصل الثامن: ويتطرق إلى قضايا الارتباط والانحدار بين المتحولات وكيفية حساب معادلة الانحدار واستخدامها في عمليات التنبؤ ثم حساب مجال الثقة للقيم المتنبأ بها.
- الفصل التاسع: ويخصص لاستعراض التحليل اللوجستي الثنائي البسيط وكيفية استخدامه في دراسة الانحدار لمتحول ثنائي مع متحول كمي مستمر أو مرتب.
- الفصل العاشر: ويتناول أهم الاختبارات اللامعلمية المستخدمة في مسائل البحث العلمي مثل: اختبار كاي مربع بأشكاله المختلفة واختبار (كولموغوروف- سميرنوف) واختبار(ليليفيوز) واختبار (غاما) ...الخ.

ولقد حاولنا أن تكون هذه المحاضرات مبسطة ومختصرة وبعيدة عن التعقيدات الرياضية، وذلك حتى تكون في متناول جميع الطلاب والباحثين العرب من مختلف الاختصاصات . علماً بأنه يمكن للباحث الملم بمبادئ علم الإحصاء أن يتجاوز الفصول الأربعة الأولى, ويختار ما يهمه من الفصول الأخرى.

وفي هذه المناسبة نتقدم بجزيل الشكر لكل السادة: الدكتور ياسر علوش والدكتور رسلان العلي والدكتورة ربا العبد الله وطالب الدكتوراة خضر العكاري على مساعدتهم لنا في تدقيقها وتنسيقها ونشرها، كما نتوجه بالشكر للأنسة سوزان صقر، التي قامت بطباعتها بكل صبر وإتقان.

ونأمل أن نكون قد وفقنا في تقديم ما يساعد طلابنا وباحثينا في جميع البلدان العربية على إعداد بحوثهم العلمية, ويساهم في تقدم وتطور وبناء وطننا العربي الكبير.

والله من وراء القصد

المؤلفان

اللاذقية في 2021/01/15

الجـزء الأول

القضايا النظرية والمنهجية للبحث العلمي

الفصل الأول: أسس البحث العلمي

الفصل الثاني: مناهج البحث العلمي

الفصل الثالث: التحضير للبحث العلمي

الفصل الرابع: خطوات وطرائق جمع البيانات

الفصل الخامس: قواعد كتابة الرسالة أو الأطروحة

الفصل الأول
أسس البحث العلمي

1-1 : تمهيد:

إن الإنسان يبحث منذ أقدم العصور عن إجابات لتفسير الظواهر الطبيعية والاجتماعية التي تحيط به، وذلك من أجل فهم جوهرها والاستفادة من حركتها والوقاية من أخطارها والتحكم بمجاريها. وكان تفكيره في البداية يعتمد على تفاسير خيالية فينسب حركة هذه الظواهر إلى قوى مجهولة. وعندما تعمقت تجربته في الحياة راح يستخدم المحاكمات العقلية، ويضع أجوبة فلسفية لتفسير مجريات الأمور. ولكنه في النهاية انتقل بتفكيره إلى استخدام أسلوب البحث العلمي والتجريب. ومن هنا ظهر مصطلح البحث العلمي الحديث، وتحولت المعارف الإنسانية إلى علوم متخصصة في كل مجال من مجالات الحياة والطبيعة.

ومع إن أسلوب البحث العلمي يختلف من علم لآخر إلا أن هيكله العام يأخذ شكلاً موحداً في جميع العلوم ويعتمد على الأسس التالية:

1- **مقولة السببية:** وتنطلق هذه المقولة من أن وراء حدوث كل ظاهرة أسباب وعوامل أدت إلى حدوثها، وهذه المقولة تضع على عاتق الباحث مهمة البحث عن تلك الأسباب وتحديد مدى علاقتها بالظاهرة المدروسة. كما أن هذه المقولة تلزم الباحث باتباع أسلوب البحث التجريبي وتبعده عن التفسيرات الغيبية المختلفة. ومن المفيد أن نذكر القارئ بأن اعتماد هذه المقولة في البحوث العلمية هو الذي أدى إلى اكتشاف معظم القوانين العلمية وإلى تفسير مختلف الظواهر الطبيعية وإلى تحقيق كل المنجزات والمكتشفات الحالية.

2- **مبدأ الأمانة والدقة:** وهو مبدأ أساسي في البحث العلمي، وأن الأمانة والدقة هما صفتان أساسيتان يجب أن يتمتع بهما ويحافظ عليهما كل باحث، حيث أنهما يفرضان عليه أن يقوم بتسجيل مختلف قياساته وملاحظاته بأمانة كاملة ودقة بالغة، دون أي تحيز أو تحريف، لأن أي تحريف أو خطأ في تلك القياسات أو الملاحظات سيؤدي إلى نتائج مغلوطة، كما أن الأمانة تقتضي من الباحث أن يشير ويعترف بجميع الحقائق، التي اكتشفها الآخرون قبله دون أي تحوير أو تزوير، ولا يجوز أن ينسب إلى نفسه ما انجزه الآخرون قبله.

3- مبدأ المقارنة: وهو يقتضي أن يقوم الباحث بمعالجة البيانات التي حصل عليها ومقارنتها مع االبيانات المماثلة لها في الدراسات السابقة، وأن يستخلص منها الفروقات والنتائج الممكنة . ولكن عملية المقارنة هنا يجب أن تتم ضمن الشروط التالية: أن تكون المتحولات الخاضعة للمقارنة متماثلة

بطبيعتها وتعاريفها، وأن يكون لبياناتها المتقابلة واحدة قياس موحدة، وأن تكون طريقة معالجتها قبل المقارنة موحدة، وان تكون الظروف المحيطة موحدة أو متشابهة على الأقل. وإن أي خلل في هذه الشروط يؤدي إلى استنتاجات خاطئة.

4- **مبدأ الشمولية:** وهذا يعني أنه على الباحث أن يقوم بدراسة جميع الأمور المتعلقة بالموضوع المدروس وأن يتناول ذلك من جميع الجوانب، وأن يسعى في النهاية للحصول على نتائج تتسم بالشمولية وتنطبق على جميع الظواهر أو المواد المماثلة. وكمثال على ذلك نذكر أن جميع القوانين التي اكتشفها الإنسان تتصف بالشمولية وهي تطبق على جميع الظواهر والمواد المماثلة: كقانون أرخميدس للأجسام المغمورة في الماء أو غيره، وقانون الجاذبية الأرضية وقانون التمدد الحراري ...الخ.

وأخيراً لابد من أن نشير إلى أن مبدأ الشمولية لا يعني التوصل إلى حقيقة مطلقة بل إلى حقيقة نسبية مشروطة بظروف الدراسة ووسائلها، وليست مستقلة عن الحقائق الأخرى.

5- **مبدأ التنظيم:** تختلف أساليب البحث العلمي عن أساليب التفكير العادي البسيط بأنها أساليب منظمة وتستند إلى مناهج ومفاهيم ونظريات علمية، وبرامج محددة وفيها يتم اختبار الفروض الابتدائية بوسائل دقيقة وعبر مراحل متتالية، وبالتالي يتم التوصل إلى نتائج محددة.

1-2 : مفهوم البحث العلمي:

إن الباحثين لا يتفقون على تعريف موحّد للبحث العلمي وذلك بسبب تعدد الأساليب العلمية المستخدمة في البحوث المختلفة، ولكن معظم الآراء تتفق على أن مفهوم البحث العلمي يشمل القضايا التالية:

- إن البحث العلمي هو عملية دقيقة ومنظمة لإيجاد حلول لمختلف المشكلات التي تواجهها الإنسانية.
- إن البحث العلمي يهدف إلى توسيع دائرة معرفة الإنسان من أجل تعزيز قدرته على التكيف مع البيئة المحيطة به والتحكم بنشاطها واكتشاف العلاقات بين أطرافها.
- إن البحث العلمي يعتمد على اختبار الفرضيات الابتدائية ولا يعلن النتائج إلا بعد فحصها والتأكد من صحتها تجريبياً.
- إن البحث العلمي يتناول جميع ميادين الحياة ويدرس جميع المشكلات ويطبق في جميع المجالات المعرفية.

ونتيجة لهذه المفاهيم يمكننا تعريف البحث العلمي بما يلي:

البحث العلمي: هو مجموعة من الجهود المنظمة والاستقصاءات المبرمجة التي تعتمد على التجربة والمنطق، وتهدف إلى إيجاد حلول مناسبة لمختلف المشكلات التي تواجه الإنسانية، وإلى توسيع دائرة معارف الإنسان وتعزيز قدراته العامة على التكيف مع البيئة المحيطة به والتحكم بها والاستفادة من خبراتها.

3-1 : أهمية البحث العلمي:

إن أهمية البحث العلمي تتبع من أن الإمكانات والوسائل العلمية المتبعة في البحث العلمي تتمتع بثقة جميع الباحثين والمهتمين، وإن أساليبه ووسائله ترفع من قدرة الإنسان على فهم مشكلاته وتساعده على إيجاد الحلول المناسبة لها، كما إنها تمهد له الطريق لدراسة الأبحاث العلمية التي يقوم بها الآخرون وبالتالي تمكنه من الحكم على مدى دقة وجدية تلك الأبحاث والدراسات.

وهكذا نجد أن أساليب البحث العلمي أصبحت وسيلة حتمية لإجراء أي نشاط فكري أو تجريبي، وهذا ما يفرض على الباحثين اتقان عدد من المهارات الأساسية لعمليات التصميم والقياس والتسجيل والنقل والتحليل والعرض والاستقراء والاستنتاج وحساب الدقة والثقة والتنبؤ بحركات الظواهر واستخلاص النتائج واقتراح الحلول ...الخ. وإن اتقان هذه المهارات يحتاج إلى اتقان العمل على العديد من الأجهزة والأدوات المتطورة ومعرفة ضبط العوامل والمتغيرات، وبرمجة المدخلات والمخرجات، وتحليل البيانات واستخلاص النتائج ...الخ.

وإن اتباعنا لأساليب البحث العلمي في دراسة مشكلاتنا الحياتية يحسن بشكل ملحوظ أساليب عملنا ويطور تفكيرنا، وهذا ما يؤدي لزيادة في الإنتاج القومي ورفع مستوى حياتنا الشخصية والاجتماعية والاقتصادية.

4-1 : خصائص البحث العلمي:

مما تقدم يمكننا أن نستخلص بعض الخصائص التي تتميز بها أساليب البحث العلمي وهي:

1- **انسجام النتائج:** إن النتائج التي نحصل عليها بواسطة أساليب البحث العلمي وضمن نفس الظروف لا يمكن أن تكون متناقضة، أي لا يمكن بواسطتها إثبات الشيء ونقيضه في الوقت نفسه، ففي تجربة تحليل الماء إلى عنصريه (الأوكسجين والهيدروجين) لا يمكننا التوصل إلى نتيجة تؤكد وجود الأوكسجين في الماء وأخرى تنفي وجوده في الوقت نفسه .

2- **تراكم المعرفة:** إن المعرفة الإنسانية في لحظة ما هي حصيلة تراكم جميع المعارف الإنسانية السابقة، فالعلماء يبنون نظرياتهم واختراعاتهم بناءً عمودياً وأفقياً، وينطلق كل عالم من نهاية ما توصل إليه غيره . فالنظريات العلمية الحديثة تكمل أو توسع أو تعمم النظريات العلمية السابقة، فنظريات الهندسة الفراغية هي تعميم لنظريات الهندسة المستوية، والأعداد المركبة هي توسع للأعداد الحقيقية, والتكامل هو نهاية المجموع، والمصفوفات هي تعميم آخر لمفاهيم الأشعة والأعداد المركبة ...الخ.

3- **الاعتماد على التجربة والمنطق:** إن أسلوب إجراء التجارب لإثبات أو اكتشاف الحقائق الطبيعية والاجتماعية هو أحد أصول البحث العلمي، فالتجربة هي الوسيلة الأولى لتأكيد أو رفض الفرضيات الموضوعية حول أية ظاهرة. كما أن المحاكمة العقلية المستندة على أسس المنطق السليم تعد من أهم

وسائل التفكير والتحليل في فهم الموضوع المدروس والعوامل المؤثرة به وفي تحليل معطيات التجارب واستخلاص النتائج المرجوة منها .

5-1 : ميادين البحث العلمي:

يمكن تصنيف ميادين البحث العلمي إلى مجموعتين رئيستين هما: ميادين الظواهر الطبيعية وميادين الظواهر الاجتماعية .

1- **ميادين الظواهر الطبيعية:** وتشمل جميع الظواهر الطبيعية مثل قضايا الفيزياء والكيمياء، والأرض والفضاء، والطب والدواء، والماء والهواء، والنبات والحيوان ...الخ.

ولقد حقق الإنسان إنجازات هائلة في هذه الميادين لأنه أخضع ظواهرها للتجربة والخطأ دون أن يلاقي أية اعتراضات أو تحفظات خارجية. كما إن الطبيعة المادية للظواهر الطبيعية مكنته من تحديد عناصرها وحصر تغيراتها واخضاع نشاطاتها للتجربة والقياس، وبالتالي استطاع أن يقوم باستخلاص أهم خصائصها وتوظيفها في خدمته العامة والخاصة.

2- **ميادين الظواهر الاجتماعية:** وتشمل جميع الظواهر المتعلقة بالحياة الإنسانية مثل: المشكلات المرضية والنفسية والتربوية والاقتصادية ...الخ.

وهذه المشكلات مرتبطة بنظم القيم والعادات والتقاليد الاجتماعية، ومن الصعب على الباحث اقتحامها وإجراء التجارب على أصحابها دون معاناة أو مخاطرة, لأنه لا يمكن للباحث أن يخضع الإنسان أو المجتمع لتجارب تضر بصحته أو بمصالحه أو تتعارض مع قيمه وتقاليده، وإذا تمكن من التغلب على هذه المخاطر يصطدم الباحث بصعوبة تحديد عناصر تلك الظواهر وقياس تغيراتها وضبط نشاطاتها ...الخ.

ورغم هذه المخاطر والمحاذير فإن أساليب البحث العلمي بدأت تنتشر بسرعة هائلة في هذه الميادين، فأُنشئت المخابر النفسية والتربوية وأُحدثت مراكز البحوث الطبية، وتطورت وسائل المعلوماتية وأخذت تساهم في دراسة المشكلات الاجتماعية والاقتصادية وتفسير أسبابها وإيجاد الحلول المناسبة لها. ونتيجة لذلك أصبح البحث العلمي في جميع المجتمعات المتطورة منهجاً معتمداً في دراسة المشكلات الاجتماعية ودخلت أساليبه مختلف شؤون الحياة وتمكنت من تحقيق إنجازات كبيرة في هذه المجالات.

ولكن البحث العلمي في الميادين الاجتماعية مازال يصطدم بكثير من المصاعب والعراقيل وخاصة في المجتمعات المتخلفة، وإن أبرز هذه العوائق تتجسد في الجوانب التالية:

- انتشار الفكر الإسطوري أوالخرافي، الذي يضع تفسيرات خيالية للظواهر الاجتماعية والطبيعية ويربطها بأمور غير علمية: كالسحر والتنجيم وتحضير الأرواح وحركة الأبراج وغيرها من الخرافات الشائعة في المجتمعات المتخلفة.

- الالتزام بالأفكار المسيطرة في المجتمع, وهذا ما يجعل الإنسان البسيط يستسلم لبعض الأفكار والمقولات الشائعة لدى الناس منذ القدم اعتقاداً منه أن هذه الأفكار أو النظريات ما كانت لتبقى لو لم تكن صحيحة، وذلك رغم أن التجارب العلمية الحديثة برهنت بشكل قاطع على بطلانها. وكمثال على ذلك نذكر ما كان يعتقده الناس حول عدم كروية الأرض أو دورانها حول الشمس، وما كانوا يقدمونه من تفسيرات لحادثتي الكسوف والخسوف ...الخ.
- الشك بقدرة العقل البشري: وهذا يعني أن بعض الناس يعتبرون أن العقل البشري عاجز عن الوصول إلى الحقيقة المطلوبة ولا يصلح لقيادة الشؤون الإنسانية ولا لفهم وتفسير أسرار الكون، ويستسلمون لبعض النظريات الجاهزة .

إن هذه العوائق وغيرها سرعان ما تنحسر وتتبدد عندما ينطلق البحث العلمي في دراسة الظواهر المختلفة وإيجاد التفسيرات العلمية لها. وإن الإنجازات العلمية الحديثة كشفت بطلان كل الأفكار المغلوطة ووضعت المجتمعات المتخلفة أمام حقائق جديدة، ودفعتها للتخلص من أوهامها القديمة، ودعتها لاتباع أساليب البحث العلمي واستخدام أدواته وأجهزته الحديثة.

1-6 : مناهج البحث العلمي:

يختلف الباحثون حول مفهوم كلمة منهج، وبدون التعرض إلى تلك الاختلافات سنميز المناهج والأساليب التالية:

1- المنهج الاستقرائي: ويستخدم في الانتقال من القواعد الخاصة إلى القانون العام.
2- المنهج الاستنتاجي: ويستخدم في الحصول على نتائج خاصة من القانون العام.
3- المنهج التاريخي: ويستخدم في دراسة وتتبع تطور الأحداث والأحوال عبر الزمن.
4- المنهج التجريبي: ويستخدم في إعداد الدراسات الاجتماعية والبحوث الطبيعية ويعتمد على إجراء تجارب اصطناعية على الظاهرة المدروسة.
5- المنهج الوصفي التحليلي: ويستخدم في جميع البحوث الطبيعية والاجتماعية، ويعتمد على الوصف والتحليل والمقارنة لمختلف المؤشرات الكمية والنوعية المحددة للظاهرة المدروسة، والتي تم الحصول عليها من الميدان أو من نتائج التجارب الاصطناعية.
6- منهج المسح الإحصائي: ويستخدم في البحوث الميدانية التي تتناول الظواهر الطبيعية والاجتماعية، ويعتبر هذا المنهج مرافقاً للمنهجين الوصفي والتجريبي.
7- منهج تحليل المضمون: ويستخدم في دراسة مضمون ظاهرة معينة لإبراز سماتها الأساسية .
8- منهج دراسة الحالة: ويستخدم في دراسة حالة معينة لإظهار إيجابياتها وسلبياتها.
9- منهج تحليل النظم: ويستخدم في دراسة النظم الاجتماعية والطبيعية والاصطناعية ومعالجة مشكلاتها.

وسنخصص فصلاً كاملاً لشرح هذه المناهج.

7-1 : مراحل البحث العلمي: تمر عملية البحث العلمي بعدة مراحل أساسية هي:

1- **مرحلة التحضير للبحث العلمي:** وتشمل الأمور التالية:

- اختيار موضوع البحث.
- مراجعة الدراسات والمراجع المتعلقة بالموضوع المختار.
- تلخيص الدراسات السابقة وإبراز مدى اختلاف البحث عنها.
- تحديد مشكلة البحث وصياغتها بدقة.
- تحديد أهمية البحث وأصالته.
- تحديد الهدف من البحث.
- تحديد وتعريف المصطلحات الجديدة.
- تحديد وفرز المتحولات والتوابع المستخدمة في البحث.
- تحديد المجال المكاني والزماني للبحث.
- تحديد مجتمع وعينة البحث.
- وضع فرضيات البحث وصياغتها بطريقة علمية.
- تحديد منهج وأدوات البحث.
- إعداد مخطط البحث.
- إعداد قائمة المراجع والمصادر.
- تقديم البحث إلى الجهات المعنية لاعتماده وتسجيله.

2- **مرحلة جمع البيانات:** وتشمل الأمور التالية:

- تحديد طريقة جمع البيانات (القياس، الاستخلاص، الملاحظة، المقابلة، الاستبيان) .
- وضع خطة ولوازم جمع البيانات.
- تصميم الاستثمارات والجداول اللازمة.
- تدريب الكوادر المشاركة.
- تسجيل البيانات اللازمة على الاستمارات أو الجداول المعدة لذلك بعد تدقيقها.

3- **مرحلة معالجة وتحليل البيانات:** تتم معالجة البيانات حسب هدف الدراسة وتشمل الأمور الأساسية التالية:

- ادخال البيانات على الحاسوب حسب البرنامج المستخدم SSPS أو SAS أو غيرهما.
- ترتيب أو تبويب البيانات حسب الهدف.
- حساب النسب والمعدلات والمتوسطات والانحرافات المعيارية.
- حساب المؤشرات الوصفية لهذه البيانات.

- تقدير المؤشرات المدروسة في المجتمع.

4- **اجراء الاختبارات اللازمة للتحقق من الفرضيات:**

- اختبار الفرضيات البحثية المذكورة في مقدمة البحث.
- دراسة الارتباط والانحدار بين المتحولات وإجراء بعض التنبؤات.
- تطبيق بعض الأساليب الأخرى إن لزم ذلك وحسب الحاجة.

5- **كتابة التقرير النهائي:** وتشمل الأمور التالية:

- صياغة مقدمة البحث: وتضم مشكلة البحث وأهمية وأهدافه والمجال المكاني والزماني له، والمنهج المستخدم فيه وتعريف المصطلحات والمتحولات وصياغة الفرضيات ووسائل اختبارها ...الخ.
- كتابة الفصول حسب المخطط الموضوع.
- استخلاص النتائج من التحليل والاختبارات والتطبيقات وتقديم المقترحات والتوصيات.
- إعداد قائمة المراجع والمصادر حسب الأصول.
- وضع فهرس لمحتويات البحث.

وسنخصص فصلاً كاملاً لكل من هذه المراحل.

الفصل الثاني

مناهج البحث العلمي

سنتعرض في هذا الفصل إلى المفاهيم الأساسية لأهم مناهج البحث العلمي، تاركين للباحث أمر التعمق في المنهج الذي يثير اهتمامه ويخدم البحث الذي يقوم به , وهذه المناهج هي:

2-1 : المنهج الاستقرائي:

وهو من أهم مناهج البحث العلمي ويستخدم في البراهين الرياضية وصياغة العلاقات الفيزيائية والكيميائية وغيرها، ويعتمد على مبدأ الانتقال من القواعد الخاصة إلى القانون العام. فمثلاً عند دراسة تأثير الحرارة على الحديد يقوم الباحث بإجراء تجارب متعددة وبدرجات حرارة مختلفة فيلاحظ أن الحديد يتمدد نتيجة للحرارة وإن تمدده يزداد مع ازدياد درجة الحرارة (في حدود معينة)، فيستخلص الباحث أن الحديد يتمدد بالحرارة وإن تمدده يتناسب طرداً مع درجة الحرارة. ويمكن للباحث أن يعيد التجربة على جميع المعادن فيجد أنها تتمدد بالحرارة، ولكنه سيلاحظ أن تمددها يختلف من معدن لآخر، وهكذا يستخلص الباحث نتيجة هامة وهي " إنّ المعادن تتمدد بالحرارة وإن تمددها يتناسب طرداً مع زيادة الحرارة وإن ذلك التمدد يختلف من معدن لآخر" . ثم يقوم بصياغة معادلات التمدد لكل معدن أو بصياغة معادلة موحدة لتمدد المعادن.

وباستخدام هذا المنهج تم استنباط الكثير من القوانين الرياضية والفيزيائية. وكمثال على ذلك نستعرض كيف تم استخلاص قانون الفائدة المركبة .

نفترض أن شخصاً أودع في المصرف مبلغاً قدره S_0 ل.س وبفائدة مركبة سعرها P سنوياً. وأراد أن يحسب حصيلة مبلغه بعد n سنة. ولمعالجة هذا الأمر قام بحساب الحصيلة سنة بعد سنة أخرى كما يلي:

إن حصيلة المبلغ بعد مرور سنة واحدة تساوي المبلغ S_0 مضافاً إليه مقدار الفائدة المستحقة والبالغ $P * S_0$، أي أن:

$$S_1 = S_0 + P * S_0 = S_0(1 + P)$$

وإن حصيلة المبلغ بعد مرور سنتين تساوي S_1 مضافاً إليه مقدار الفائدة الجديدة على S_1، أي أن:

$$S_2 = S_1 + P * S_1 = S_1(1 + P) = S_0(1 + P)(1 + P) = S_0(1 + P)^2$$

وإن حصيلة المبلغ بعد مرور ثلاث سنوات تساوي:

$$S_3 = S_2 + P * S_2 = S_2(1 + P) = S_0(1 + P)^3$$

وبمتابعة عملية الاستقراء إلى n سنة حصل على القانون العام التالي:

$$S_n = S_0(1 + P)^n$$

ورغم إن هذا المنهج يمكّن الباحث من التوصل إلى نتائج مؤكدة ويتمتع بمصداقية كبيرة، إلا إنه يشترط أن يتناول الباحث جميع الجزئيات المتعلقة بالموضوع المدروس وضمن نفس الشروط أو الظروف المحيطة. وبما أن هذا الأمر غالباً ما يكون مكلفاً أو متعذراً، لذلك يتم اللجوء إلى ما يسمى بمنهج الاستقراء الناقص، وهو الاستقراء الذي يعتمد على دراسة عينة محدودة من عناصر الموضوع المدروس، وبناء على معطياتها يتم تعميم النتيجة واستخلاص القانون العام. وتكون درجة الثقة بهذه النتيجة مرتبطة بحجم العينة وبدرجة تمثيلها للمجتمع المدروس.

2-2 : المنهج الاستنتاجي:

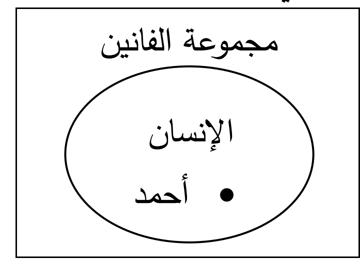
وهو منهج يعتمد على استخلاص بعض الأمور الخاصة من قواعد عامة، مستنداً إلى أسلوب المنطق الرياضي الذي يشترط وجود مقدمة كبرى ومقدمة صغرى، ينتج عنهما نتيجة محددة، وحتى تكون تلك النتيجة صحيحة يجب أن تكون المقدمتان صحيحتين، وأن تكون المقدمة الصغرى محتواة في المقدمة الكبرى، وأن يكون الطرف الأول في كلتيهما محتوى في الطرف الثاني .

وكمثال على ذلك نأخذ القضية التالية:

المقدمة الكبرى : كل إنسان فان

المقدمة الصغرى : أحمد إنسان

النتيجة : أحمد فان



وبما أن المقدمتين السابقتين صحيحتان، والصغرى محتواة في الكبرى، والطرف الأول في كلتيهما محتوى في الطرف الثاني، فإننا حصلنا على نتيجة صحيحة.

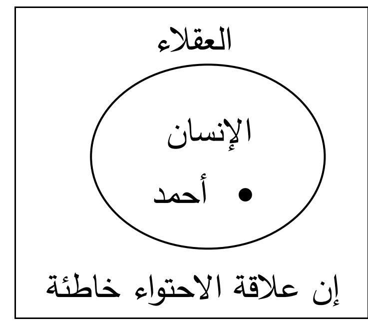
ولكن الأمر لا يتم دائماً بهذه السهولة، فإذا كانت إحدى المقدمتين خاطئة (كلياً أو جزئياً)، فإننا نحصل على نتيجة خاطئة (كلياً أو جزئياً).

وكمثال على ذلك نورد القضية التالية:

المقدمة الكبرى : كل إنسان عاقل ؟

المقدمة الصغرى : أحمد إنسان

النتيجة (؟) : أحمد عاقل (؟)

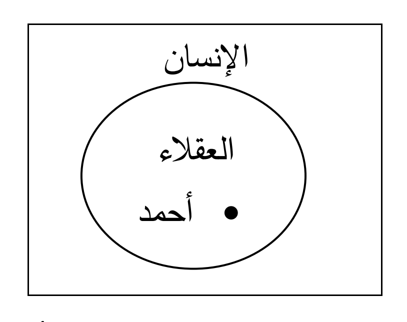


وهي نتيجة ليست مؤكدة، لأن أحمد قد يكون غير عاقل . ولقد حصلنا على هذه النتيجة لأن المقدمة الكبرى خاطئة جزئياً (طرفها الأول غير محتوى في طرفها الثاني) . وكان يجب وضع القضية على الشكل التالي:

المقدمة الكبرى : كل عاقل إنسان

المقدمة الصغرى : أحمد عاقل

النتيجة : أحمد إنسان



إن الأسلوب الذي يعتمد عليه هذا المنهج يشبه خاصة التعدي في الجبر, والتي تنص على أنه إذا كان:

$$a < b \quad , \quad b < c \quad \Rightarrow \quad a < c$$

وهكذا نجد أن هذا المنهج يعتمد على علاقة الانتماء والاحتواء في المجموعات, والتي يمكن كتابتها كما يلي:

إذا كان العنصر X ينتمي إلى المجموعة A , وكانت المجموعة A محتواة في المجموعة B : فأن العنصر X ينتمي أيضاً إلى المجموعة B .

ومما يؤخذ على المنهج الاستنتاجي إنه يعرّض الباحث للخطأ إذا كانت إحدى المقدمتين خاطئة (كلياً أو جزئياً). لذلك ننصح الباحثين بتحري الدقة عند استخدام هذا المنهج في عمليات البحث, ولايفعلوا كما فعل حفيدي ليقنعني بأن (الذئب يأكل العشب) فقال: الذئب يأكل الخروف, والخروف يأكل العشب ،إذن : الذئب يأكل العشب(لقد حصل على هذه النتيجة الخاطئة لأن العلاقة بين طرفي كل مقدمة ليست علاقة احتواء بل علاقة أكل).

2-3 : المنهج التاريخي:

يستخدم هذا المنهج في دراسة تطور الظواهر كمياً ونوعياً عبر الزمن الماضي، وبما ان تلك الدراسات تتناول ظواهر من الماضي فإن الباحث لا يستطيع استرجاع ذلك الماضي وحصر العوامل المؤثرة بظواهره، لذلك يلجأ إلى مصادر خاصة لتزويده بالمعلومات اللازمة، وأهم هذه المصادر هي:

- السجلات الإحصائية والمدنية والعقارية.
- الوثائق الرسمية والاتفاقيات الحكومية والرسائل المتبادلة.
- الصحف والكتب والمجلات العلمية والأدبية ووسائل الإعلام الأخرى.
- الآثار المادية المتوفرة في المكان أو المنتشرة في جميع أنحاء العالم.
- الكتابات والروايات التاريخية والقصص الأدبية والأعمال الفنية.
- المذكرات والسير الذاتية للكتّاب والسياسيين.
- الشهود الأحياء وأقوالهم الشفوية أو شهاداتهم المدونة.
- الدراسات السابقة التي تتناول الموضوع المدروس.

ولكن هذه المصادر قد تكون عرضة للتعديل أو التزوير أو الاهمال أو النسيان أو المبالغة أو الاختزال. ولهذا فإنه على الباحث أن يقوم بعملية نقد شاملة خارجية وداخلية لمحتوى تلك المصادر للوقوف على

الحقائق الثابتة فيها، ولا يستطيع الباحث أن يقوم بذلك إلا إذا توفرت لديه معرفة واسعة بالتاريخ واللغات والقيم والنظم والعادات والتقاليد الاجتماعية التي كانت سائدة في تلك المرحلة.

إن الصعوبات التي تعترض هذا المنهج قد تضعف الثقة في نتائجه، إلا أن اتباع الأسلوب الموضوعي في البحث وعدم التحيز لبعض مجريات الأمور التاريخية يؤمنان الظروف المطلوبة للوصول إلى نتائج علمية هامة.

وتبرز أهمية المنهج التاريخي في معالجة المشكلات المتعلقة في الأمور التالية:

- الكشف عن الأصول الحقيقية للنظريات والمبادئ العلمية.
- الكشف عن المشكلات التي كان يعاني منها الإنسان في الماضي والتعرف على أساليب معالجتها.
- تحديد العلاقات التي تربط بين الظواهر الاجتماعية والاقتصادية والثقافية والبيئية وغيره.
- التنبؤ بأحوال وقيم الظواهر المدروسة وتقدير النتائج المترتبة على ذلك.

2-4 : المنهج التجريبي:

وهو منهج هام يطبق في جميع ميادين البحث العلمي، ويعتمد على إجراء تجارب اصطناعية على الظاهرة المدروسة وعلى إسلوب المقارنة بين نتائجها، ويقوم هذا المنهج على مبدأ أساسي هو:

إذا كان لدينا موقفان متشابهان A وB وأدخلنا على الموقف A (أو حذفنا عنه) عاملاً جديداً x فإن الفروقات الناتجة بين A وB تعود إلى تأثير (أو حذف) العامل x.

ولتطبيق هذا المنهج يجب على الباحث أن يقوم باتخاذ الإجراءات التالية:

1- تحديد جميع العوامل المؤثرة في الظاهرة المدروسة.

2- تحديد العامل المستقلx (المتحول التجريبي) الذي نريد دراسة مدى تأثيره على الظاهرة المدروسة.

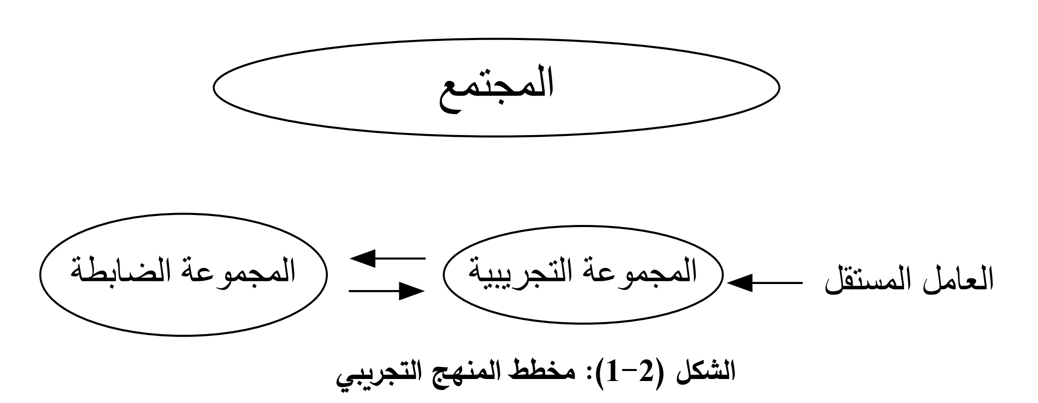
3- تحديد العامل التابع (المتحول الناتج) الذي ينتج عن تأثير العامل المستقل على الظاهرة.

4- تثبيت قيم العوامل الأخرى المؤثرة في العامل التابع – ماعدا قيم العامل المستقل- وذلك لتبيان مدى تأثير العامل المستقل على العامل الناتج بشكل منفرد.

5- تشكيل مجموعتين متكافئتين لتطبيق إجراءات التجربة على أحدهما ثم مقارنة تغيراتها مع الأخرى. ويطلق على هاتين المجموعتين المصطلحين التاليين:

- المجموعة التجريبية: وهي المجموعة التي ستخضع للتجربة وستتعرض لتأثيرات العامل المستقل مع تثبيت العوامل الأخرى. وتجرى التجربة على أساس التحكم التدريجي في تغيرات العامل المستقل وقياس تأثيراتها المقابلة.
- المجموعة الضابطة: وهي مجموعة خاصة تدخل في التجربة دون التأثير عليها بالعامل المستقل ، أي أنها تبقى تحت الظروف العادية، التي تشترط ثبات جميع العوامل الأخرى بما فيها العامل المستقل، وتستخدم المجموعة الضابطة من أجل مقارنة نتائج التجربة في

المجموعة التجريبية مع النتائج المقابلة لها في المجموعة الضابطة، والشكل التالي يوضح ذلك.



الشكل (2-1): مخطط المنهج التجريبي

- **أساليب تصميم التجارب:**

تستخدم في هذا المنهج عدة أساليب لتصميم التجارب، أهمها الأساليب التالية:

1- أسلوب المجموعة الواحدة: نظراً لصعوبة تأمين مجموعتين متكافئتين تماماً، يقوم الباحث بتطبيق تجربته على مجموعة واحدة وذلك باتباع الإجراءات التالية:

- تحديد وقياس قيم المتحولات المستقلة والتابعة قبل إجراء التجربة وتسجيلها بدقة. وتسمى نتائج هذا القياس بنتائج الاختبار القبلي (قبل إجراء التجربة)، وهي تمثل نتائج المجموعة الضابطة.
- القيام بإجراء التجربة المبرمجة وتطبيق تغيرات العامل المستقل بشكل تدريجي على نفس المجموعة باعتبارها مجموعة تجريبية.
- تسجيل النتائج التي يتم الحصول عليها تحت اسم نتائج الاختبار البعدي (بعد إجراء التجربة).
- مقارنة نتائج الاختبار البعدي مع نتائج الاختبار القبلي ومعالجتها ثم استخلاص النتائج اللازمة.

ولكن مما يؤخذ على أسلوب المجموعة الواحدة في البحوث الاجتماعية هو أن أفرادها قد ينتبهون إلى أنهم معرضون لإجراء التجربة فيعدلون من إجاباتهم وتصرفاتهم، إلا أن هذا الأمر غير وارد في البحوث الطبيعية.

2- أسلوب المجموعتين المتكافئتين: يعتمد هذا الأسلوب على اختيار مجموعتين مكافئتين لتكون إحداهما المجموعة التجريبية والأخرى المجموعة الضابطة. وحتى يضمن الباحث تأمين التكافؤ بين هاتين المجموعتين يقوم بتشكيلهما بإحدى الطريقتين التاليتين:

- طريقة السحب العشوائي: وفيها يتم سحب أفراد كل مجموعة بشكل عشوائي من المجتمع المدروس حسب الحجم المطلوب، وبعد السحب يقوم الباحث بالتأكيد من عناصر التكافؤ بين المجموعتين وذلك من خلال دراسة ومقارنة متوسطات وانحرافات ونسب بعض المؤشرات الهامة فيهما، وبعدها يحدد الباحث دون أي تحيز (بالقرعة مثلاً) أي المجموعتين ستكون المجموعة التجريبية وبالتالي تكون الأخرى المجموعة الضابطة.

- طريقة الأزواج أو التوائم: وفيها يقوم الباحث باختيار أزواج أو توائم متماثلة من حيث العوامل المؤثرة، ثم يقوم بفرز عشوائي لعنصري كل زوج ليضع الأول في المجموعة التجريبية والأخر في المجموعة الضابطة، ويكرر ذلك حسب الحجم المطلوب لكلا المجوعتين.

وحسب هذا الأسلوب يقوم الباحث بتنفيذ الإجراءات التالية:

- تحديد وقياس المتحولات المستقلة والتابعة في كلتا المجموعتين التجريبية والضابطة والتأكد من تكافئهما من حيث العوامل المؤثرة خلال التجربة المقبلة.
- القيام بإجراء التجربة وضبط جميع العوامل المؤثرة فيها، عدا العامل المستقل التجريبي الذي يعطى قيماً مدروسة ومتدرجة ويطبق تأثيره على عناصر المجموعة التجريبية.
- تسجيل نتائج تأثير العامل التجريبي على عناصر المجموعة التجريبية وذلك باستخدام الأدوات والمقاييس المناسبة.
- تسجيل قياس وتسجيل قيم المتحولات المستقلة والتابعة في عناصر المجموعة الضابطة باستخدام نفس الأدوات والمقاييس السابقة.
- مقارنة نتائج المجموعة التجريبية مع نتائج المجموعة الضابطة ومعالجتها ثم استخلاص النتائج الممكنة.

3- أسلوب المجموعات الثلاثة: ويطبق هذا الأسلوب عندما يرغب الباحث بدراسة تأثير عاملين مستقلين على الظاهرة المدروسة . وهنا يقوم الباحث أولاً باختبار تكافؤ المجموعات الثلاثة والتأكد من تحققه، ثم يحدد عشوائياً مجموعتين تجريبيّتين ومجموعة ثالثة ضابطة . ثم يقوم بتطبيق تأثير العامل المستقل الأول على المجموعة التجريبية الأولى وتأثير العامل المستقل الثاني على المجموعة التجريبية الثانية مع ضبط العوامل الأخرى في المجموعتين، وبعدها يسجل نتائج التجربتين ويقارنها مع نتائج المجموعة الضابطة ويستخلص النتائج الممكنة.

وهنا يمكن للباحث أن يقوم بإعادة التجربتين بتبديل تطبيق نفس العاملين على المجموعتين السابقتين، حيث يطبق العامل الأول على المجموعة الثانية والعامل الثاني على المجموعة الأولى، ثم يسجل نتائج هاتين التجربتين كالعادة. وبعد الانتهاء من التطبيقين المذكورين يمكن أن يقوم الباحث بإجراء المقارنات التالية:

- مقارنة نتائج تأثير العامل الأول على كلتا المجموعتين ثم مقارنتها مع المجموعة الضابطة.
- مقارنة نتائج تأثير العامل الثاني على كلتا المجموعتين ثم مقارنتها مع المجموعة الضابطة.
- مقارنة نتائج تأثير كلا العاملين المستقلين معاً على المجموعة الأولى ومقارنتها مع معطيات المجموعة الضابطة.
- مقارنة نتائج تأثير كلا العاملين المستقلين معاً على المجموعة الثانية ومقارنتها مع معطيات المجموعة الضابطة.

- **مزايا المنهج التجريبي:**

يتمتع المنهج التجريبي بالمزايا التالية:

1- يستطيع الباحث أن يعيد التجربة عدة مرات، وهذا ما يتيح له الفرصة للتأكد من صحة نتائجه ومن ثبات ومصداقية تلك النتائج.

2- يستطيع الباحث أن يتحكم في العوامل المؤثرة ويضبطها، وأن يعطي العامل التجريبي تغيرات تدريجية ويسجل تأثيراتها على الظاهرة المدروسة. وهذا الأمر يساعد الباحث على ربط النتائج بأسبابها وكشف العلاقات السببية بين أطرافها.

3- يستطيع الباحث مشاركة عدة أشخاص في إجراء التجارب وملاحظتها وتسجيل نتائجها، ولكن عليه أن يقوم بتدريبهم على إجراء التجارب المبرمجة وأن يتأكد من استعدادهم للمشاركة في العمل.

4- يعتبر المنهج التجريبي المنهج الرئيسي في كثير من مجالات البحوث الاجتماعية والنفسية والحيوية والبيئية ...الخ.

- **عيوب المنهج التجريبي:**

يعاني المنهج التجريبي من عدة عيوب هي:

1- يتطلب إجراءات إدارية وتنظيمية معقدة، وذلك من أجل الحصول على الموافقات الرسمية اللازمة وتشكيل المجموعات الاختبارية وتأمين الأمكنة والأجهزة الفنية الضرورية.

2- تتأثر نتائجه بدرجة تكافؤ المجموعات الاختبارية وبطريقة تشكيلها وبمدى تمثيلها للمجتمع المدروس وبدقة الأجهزة والمقاييس المستخدمة .

3- تتأثر نتائجه بمقدار ضبط العوامل الأخرى المؤثرة في الظاهرة، لأن عملية ضبط تلك العوامل صعبة جداً وخاصة في البحوث الاجتماعية .

4- تتأثر نتائج التجربة بالظروف الاصطناعية المرافقة لها والتي غالباً ما تختلف عن الظروف الطبيعية، وإن هذا الاختلاف يجعل الأفراد الخاضعين للتجربة يميلون إلى تعديل بعض استجاباتهم لها .

5- يواجه تطبيق المنهج التجريبي في البحوث الاجتماعية بعض الصعوبات القانونية أو الأخلاقية، كأن تمنع بعض النصوص القانونية أو تستنكر بعض القواعد الأخلاقية إجراء بعض التجارب على الإنسان أو الحيوان بدعوى إنها قد تشكل خطراً على حياته أو تؤدي إلى تغيير في سلوكه الاجتماعي أو تتعارض مع الأعراف والتقاليد.

وأخيراً نشير إلى أنه رغم هذه العيوب والصعوبات يبقى المنهج التجريبي أحد المناهج الرئيسية في البحث العلمي، وبفضل تطبيقه في مجالات العلوم الطبيعية حقق الإنسان معظم إنجازاته العلمية وأحرز أغلب اكتشافاته وانتصاراته الحديثة . وهذا ما دفع الباحثين إلى تطبيقه في مجالات العلوم الاجتماعية من أجل معالجة المشكلات التي تعاني منها المجتمعات البشرية في مختلف أنحاء العالم.

5-2 : المنهج الوصفي التحليلي:

يستخدم هذا المنهج في دراسة ووصف مختلف الظواهر الطبيعية والاجتماعية، كالظواهر الفلكية والفيزيائية والكيميائية والحيوية والاقتصادية وغيرها. ويعتمد هذا المنهج على جمع المعلومات الكمية والنوعية عن واقع الظاهرة المدروسة ثم تحليلها واستنباط أهم مواصفاتها والكشف عن العلاقات بين العوامل المؤثرة فيها والمتحولات الناتجة عنها خلال فترة الدراسة. ويمكن أن تشمل الدراسة جميع عناصر المجتمع أو يكتفي بعينة عشوائية منه.

وهناك أسلوبان لعملية الوصف هما:

1- **الوصف النوعي:** وفيه يقوم الباحث بوصف الظاهرة المدروسة وصفاً كتابياً، يحدد فيه أهم خصائصها وصفاتها النوعية وعناصرها المختلفة وكيفية تبادل التأثيرات بين أطرافها.

فلو أخذنا ظاهرة هطول الأمطار لأمكننا أن نقول: إنها ظاهرة طبيعية تحدث عندما تتواجد في السماء غيوم كثيفة وعلى ارتفاعات متوسطة وتصطدم مع تيارات هوائية باردة وتتوفر لها بقية الشروط الجوية الأخرى ...الخ.

من الملاحظة أن هذا الوصف يقتصر على الجوانب النوعية ولا يتضمن أي رقم كمي للدلالة على حجم أو كمية أي عامل من العوامل التي تدخل في تركيب الظاهرة المدروسة، وهذا ما يجعله وصفاً عاماً يثير الجدل والاختلاف.

2- **الوصف الكمي:** وفيه يقوم الباحث باستخدام أدوات أو أجهزة لقياس قيم العوامل التي تدخل في تركيب الظاهرة المدروسة . وهنا على الباحث أن يقوم بإعداد جداول خاصة لتسجيل تلك القياسات وترتيبها وتحليلها حسب ما تقتضيه أهداف الدراسة.

فالوصف الكمي لظاهرة هطول الأمطار يأخذ شكلاً آخر مثل: تهطل الأمطار عندما تتواجد في السماء غيوم كثيفة وتبلغ الرطوبة حوالي 90% وتكون على ارتفاعات تتراوح بين 2-4 كم وتصطدم مع موجات هوائية باردة تقترب درجة حرارتها من الصفر، وتتوفر بعض الشروط الجوية الأخرى ...الخ.

ويمكن للباحث أن يستخدم الأسلوبين الكمي والنوعي معاً في وصف الظاهرة المدروسة، وذلك حسب ما تسمح به الظروف وتقتضيه الحاجة. ولقد تطور مفهوم هذا المنهج فأصبح يتداخل مع المناهج الأخرى كالمنهج التجريبي ومنهج المسح الإحصائي ومع منهجي دراسة الحالة وتحليل المضمون .

6-2 : منهج المسح الإحصائي:

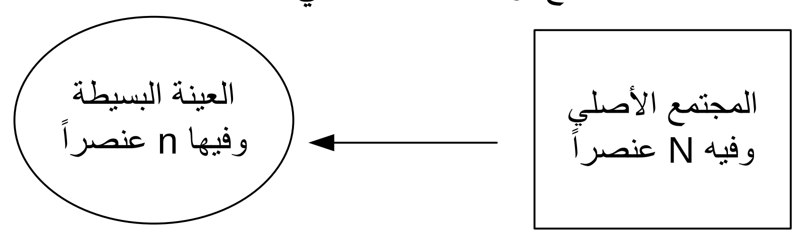
هو تطوير للمنهج الوصفي التحليلي أو شكل آخر له. ويعتمد هذا المنهج على جمع المعلومات الميدانية عن واقع الظاهرة المدروسة وتحليلها باستخدام أحدث الطرائق الإحصائية، ثم استخلاص النتائج واقتراح الحلول المناسبة. ويتميز هذا المنهج بالتعمق بدراسة الظواهر وبالإلمام بجميع العوامل المؤثرة بها وبالتركيز على جوهر القضية المدروسة. وكثيراً ما يستخدم في الدراسات الاجتماعية والسكانية والسياسية وغيرها. فهو يتناول قضايا توزع السكان وهجراتهم الداخلية والخارجية وعاداتهم وتقاليدهم الاجتماعية واتجاهاتهم وآرائهم السياسية وفئاتهم وطبقاتهم وأعمالهم وإنتاجهم ودخولهم ونفقاتهم وادخارهم وسكنهم وغذائهم وتعليمهم وصحتهم وجميع أحوالهم الأخرى. كما يستخدم هذا المنهج في الدراسات الاقتصادية المختلفة ويتناول الجوانب المادية للإنتاج والتسعير والتسويق والإدارة والتنظيم والنقل والتخزين ...الخ.

ويعتبر هذا المنهج من أهم مناهج البحث العلمي الحديث لكونه يعتمد على المعلومات الميدانية والواقعية, التي يتم جمعها عن الظاهرة باستخدام طرائق خاصة مثل: القياس والاستخلاص والملاحظة والمقابلة والاستبيان، ولأنه يستخدم في تحليلها الأساليب الإحصائية والرياضية اللازمة، وينفذها على حواسيب متطورة ومجهزة ببرامج مناسبة.

ويقوم منهج المسح الإحصائي على إجراء مسح للظاهرة المدروسة، إما في جميع واحدات المجتمع أو في عينة مسحوبة عشوائياً منه. وهناك عدة تصاميم لسحب العينات العشوائية من المجتمع أدت إلى تصنيف العينات إلى الأنواع التالية:

1- **العينة العشوائية البسيطة:** وهي العينة التي تسحب من المجتمع مباشرة ويكون احتمال سحب أي عنصر من المجتمع مساوياً لاحتمال سحب أي عنصر آخر منه، ويتحقق هذا الشرط باستخدام إحدى طرائق السحب البسيطة وهي: السحب من الصندوق أو بواسطة العجلات أو بواسطة الأرقام أو الجداول العشوائية.

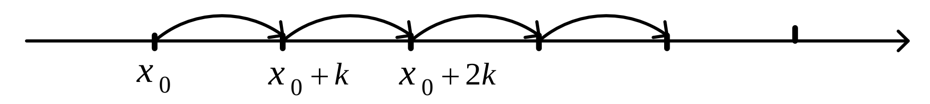
وسنرمز لعدد عناصر المجتمع بالرمز N ولعدد عناصر العينة بالرمز n , ونمثل عملية سحب العينة العشوائية البسيطة من المجتمع بواسطة الشكل التالي:



الشكل (2-2): مخطط المعاينة العشوائية البسيطة

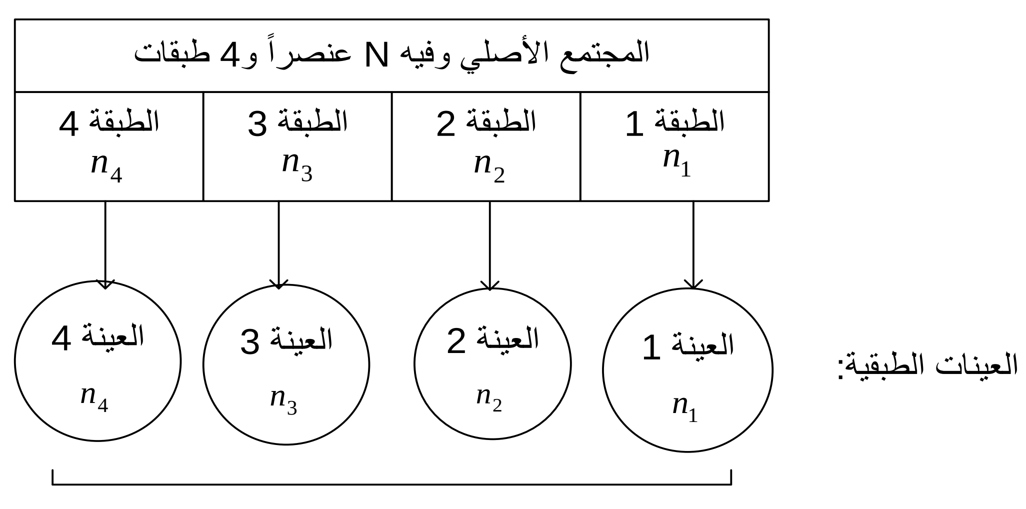
2- **العينة العشوائية المنتظمة:** وهي عينة تسحب من المجتمعات المتحركة: كرواد المعارض والمتاحف والمطاعم أو كحركة السيارات على الأوتوستراد ...الخ. ويتم سحب العينة المنتظمة باختيار عنصر البداية x_0 بطريقة عشوائية، ثم يتم اختيار العناصر الأخرى بعد مرور عدد محدد

منها (مثلاً بعد مرور كل عشرة أو عشرين عنصراً). ويمكن تمثيل سحب العينة المنتظمة بالشكل التالي :



الشكل (2-3): مخطط المعاينة العشوائية المنتظمة

3- **العينة العشوائية الطبقية:** وتستخدم عندما يكون المجتمع غير متجانس من حيث الظاهرة المدروسة، لذلك نقوم بتقسيم المجتمع إلى طبقات متجانسة، ثم نسحب عينة عشوائية بسيطة من كل طبقة وبحيث يتناسب حجمها مع حجم تلك الطبقة، فنحصل على عدة عينات بسيطة تشكل بمجموعها ما يسمى بالعينة الطبقية الكلية. ويمكن تمثيل سحب العينة الطبقية على الشكل التالي:

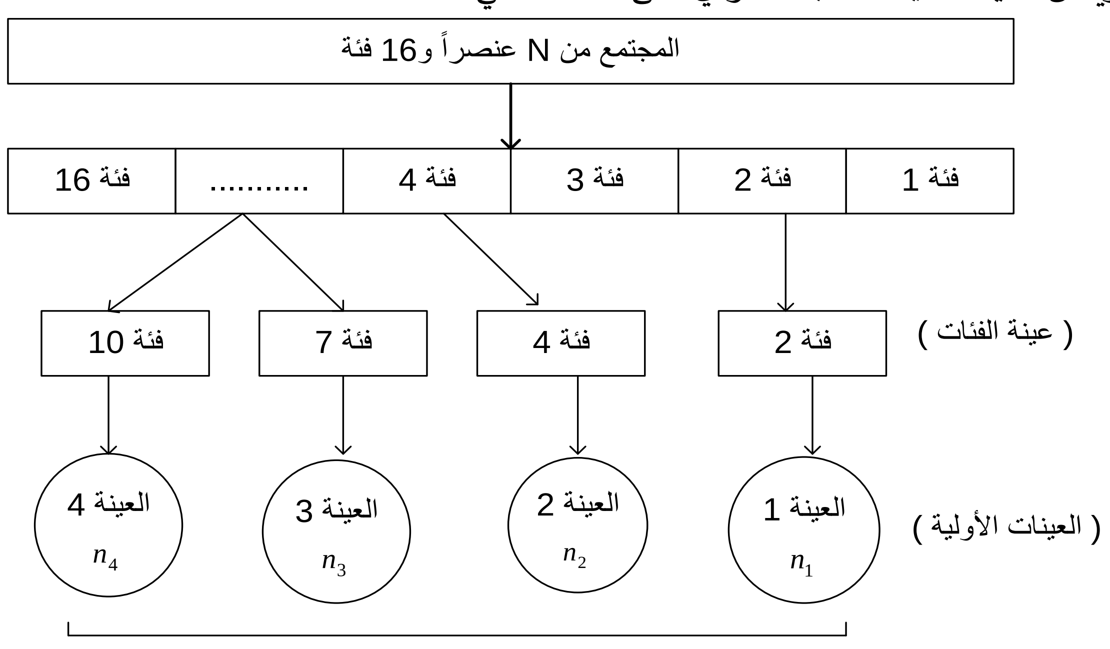


الشكل (2-4): مخطط المعاينة العشوائية الطبقية

4- **العينة العشوائية العنقودية البسيطة:** وتطبق على المجتمعات التي تكون مقسمة إلى فئات، أو يمكن تقسيمها إلى فئات كثيرة مثل: مستودعات الأدوية أو الأغذية ...الخ. وتتم عملية السحب على مرحلتين كالتالي:

- المرحلة الأولى: وفيها يتم سحب عينة من الفئات بطريقة عشوائية، وتسمى هذه العينة بعينة الفئات .
- المرحلة الثانية: وفيها يتم سحب العناصر الأولية من كل فئة مسحوبة في عينة الفئات، فنحصل على عينات أولية تشكل بمجموعها العينة العنقودية الكلية .

ويمكن تمثيل عملية السحب العنقودي على الشكل التالي:



الشكل (2-5): مخطط المعاينة العشوائية العنقودية البسيطة

2-7 : منهج تحليل المضمون:

يستخدم هذا المنهج في دراسة الإنتاج السياسي والأدبي والعلمي للمجتمعات السابقة أوالمعاصرة، كإنتاج الأفراد والجمعيات والأحزاب السياسية والحركات الإصلاحية، والنظريات الاجتماعية والإنجازات العلمية والاتجاهات الأساسية ...الخ، وذلك بهدف تحليل مضمونها والتعرف على أهم ميّزاتها واتجاهاتها وتحديد مستواها العلمي والحضاري.

وهنا يجب على الباحث أن يحدد المواد اللازمة لبحثه، وهي قد تشمل السجلات أو القوانين أو الأنظمة أو الصحف أو الكتب أو النظريات أو الاكتشافات أو الاختراعات ...الخ . ثم ينتقل إلى الدراسة المعمقة لمواد بحثه ويسجل ملاحظاته على كل فكرة واردة فيها، وبعد أن يستخلص استنتاجاته يقوم بكتابة تقريره النهائي حول البحث.

ويتميز هذا المنهج بأنه يجعل الباحث على اتصال مباشر مع مواد بحثه ويمكنه إعادة فحصها وتدقيقها والعودة إليها متى يشاء. ولكنه قد يصطدم بعدم التمكن من الاطلاع على بعض الوثائق الهامة لكونها تتسم بطابع السرية، أو يفاجأ بعدم مصداقية مواد بحثه والتي قد تكون محرّفة أو مزورة أو متحيزة أو منقوصة أو مطموسة. وإن تفادي مثل هذه الصعوبات يتطلب من الباحث التعمق في التحليل وفهم الظروف الاجتماعية السائدة وحسن اختيار مواده البحثية والقدرة على صياغة استنتاجاته دون الوقوع بأخطاء علمية كبيرة .

2-8 : منهج دراسة حالة:

يستخدم هذا المنهج لدراسة حالات متميزة، كسيرة شخص مشهور أو وضع جماعة معينة أو حالة مؤسسة قديمة. وتبدأ الدراسة بجمع المعلومات القديمة والحديثة عن تلك الحالة وتحليلها لاكتشاف أهم ميزاتها وتفسير سلوكها.

إن منهج دراسة الحالة مطبق كثيراً في الحياة اليومية والعملية، فالتاجر يدرس حالة زبونه قبل أن يشاركه، والطالب يدرس حالة زميله قبل أن يصادقه، والشاب يدرس حالة فتاته قبل أن يقدم على الاقتران منها، والفتاة تفعل ذلك قبل أن توافق على الخطوبة أو الزواج.

والطبيب يدرس حالة مريضه قبل أن يصف له العلاج، والباحث يدرس حالة المؤسسة ثم يقوم بوضع الحلول المناسبة لمعالجة مشكلاتها أو لتطوير إنتاجها أو لتسويق بضاعتها ...الخ.

ومن مزايا هذا المنهج إنه يركز اهتمام الباحث على حالة معينة ويقوم بدراستها من جميع الجوانب ويحدد علاقاتها المتشابكة مع المحيط الخارجي، ولكنه لا يسمح بتعميم نتائجه على الحالات الأخرى. إلا أنه يمكن الاستفادة من تلك النتائج في فهم ودراسة حالات أخرى تخضع لنفس ظروف الحالة المدروسة.

2-9 : منهج تحليل النظم:

يستخدم هذا المنهج في دراسة النظم الاجتماعية والاقتصادية والطبيعية . ويمكن تعريف النظام بما يلي:

النظام: هو كيان خاص يتألف من مجموعة من العناصر ترتبط فيما بينها بعلاقات منظمة وتؤدي وظائف معينة . وبذلك تكون المدرسة نظاماً، والجامعة نظاماً، والسيارة نظاماً، وجسم الإنسان نظاماً، ومؤسسة الهاتف نظاماً، ومؤسسة الكهرباء نظاماً، والكون بكامله نظاماً كبيراً يتضمن عدداً من الأنظمة الفرعية . ويتميز النظام بالخواص التالية:

1- إن له حدود معينة تعرف مكان تواجده، ويطلق على كل ما يقع خارج هذه الحدود اسم بيئة النظام، وهذه البيئة تؤثر عليه وتتأثر به.
2- إن عناصر النظام مترابطة ومتكاملة وتؤدي وظائفها ضمن ذلك الترابط والتكامل.
3- إن لكل نظام مدخلات ومخرجات، فالمدخلات قد تتألف من الناس أو الطاقة أو المعلومات أو المواد الطبيعية أو السلع الاصطناعية ...الخ. ويتم الحصول على هذه المدخلات من البيئة المحيطة بالنظام. أما المخرجات فهي المنتجات والفضلات التي يقوم النظام بتصديرها إلى البيئة المحيطة به.
4- إن النظام يقوم بعملية تحويلية للمدخلات فيخرجها على شكل منتجات وفضلات.
5- لكل نظام أهداف ووظائف، وهذا يعني أن النظام مسؤول عن إنتاج مخرجات معينة صالحة للاستخدام في الأنظمة الموجودة في البيئة المحيطة به. وهكذا تكون مخرجات نظام معين مدخلات لأنظمة أخرى، كما قد تكون مدخلاته من مخرجات بعض الأنظمة الأخرى.

6- إن النظام قد يتألف من أنظمة فرعية داخلية تعمل بالتنسيق فيما بينها لتأدية وظائف النظام بشكل كامل، فجسم الإنسان يتألف من أنظمة فرعية: كالجهاز العصبي، والجهاز الهضمي، والجهاز الدموي، والجهاز التنفسي، وغيرها.

ويعتمد تحليل النظم على النظرية الكلية (الجشتلية) في معالجة الأمور، والتي تنطلق من أن الكل أهم من الجزء أو من مجموعة أجزاء، ولا يمكن فهم الأجزاء بمعزل عن الكل، ولا يجوز المغالاة في التحليل والتخصص دون الاهتمام بعلاقات الأجزاء مع الكل وحصر دور كل منها.

وهكذا نجد أن منهج تحليل النظم ينظر إلى الموقف نظرة شاملة ويدرس جميع عناصره ويحاول أن يلمّ بجميع العوامل المؤثرة به، وهو يستند إلى الأسس التالية:

1- إن وراء كل موقف أسباباً وعوامل متعددة، ولا يجوز الاعتماد على سبب واحد في تفسير الأمور.
2- إن العوامل والأسباب المؤثرة في الموقف ليست مستقلة من بعضها بل هي متفاعلة مع بعضها البعض.
3- إن بعض العوامل المؤثرة يكون خارجياً قادماً من البيئة المحيطة وبعضها يكون داخلياً منبثقاً من النظام نفسه.

ولتطبيق هذا المنهج في دراسة النظم يجب على الباحث أن يقوم بالخطوات التالية:

1- تحليل النظام القائم ودراسة الأنظمة الفرعية فيه ومتابعة سير عملية تحويل المدخلات إلى مخرجات.
2- تحديد أهداف النظام الجديد المقترح أو المعدل من خلال أهداف الدراسة أو البحث.
3- التعرف على المشكلة وتحديدها في مختلف فروع النظام.
4- وضع الإجراءات البديلة التي تؤمن حسن سير النظام الجديد, وتعمل على التخلص من المشكلة القائمة, وتضمن عدم ظهور مشكلات جديدة، وقد تتضمن الإجراءات البديلة عدة خيارات إجرائية لمعالجة الموقف الطارئ, وفي كل لحظة يمكن اعتماد الخيار المناسب لتنفيذه على النظام أو على فروعه المختلفة.
5- تحديد أفضل الخيارات المقترحة وتطبيقه على النظام لمعالجة الموقف الطارئ.
6- تنظيم عمليات التغذية الراجعة بالمعلومات الكمية والنوعية عن سير العمل في النظام بشكل مستمر وبما يتناسب مع أهداف النظام.
7- تحليل المعلومات الراجعة ومقارنتها مع الأهداف المرسومة للنظام وتقويم الموقف ومعالجة المستجدات.
8- الاستفادة من المعلومات الراجعة في التنبؤ بمستقبل النظام لتفادي الأخطار المحتملة.

الفصل الثالث

التحضيـر للبحـث العلمـي

إن مرحلة التحضير للبحث العلمي تتألف من خطوات متداخلة ومتشابكة، ومن الصعب الحفاظ على ترتيب معين لها، وحتى لا تبقى الأمور خاضعة للصدفة والعفوية يجب على الباحث أن يكون على معرفة تامة بتلك الخطوات، وأن يقوم بتعدادها وشرحها عند كتابة تقريره النهائي، وذلك ليتمكن القارئ أو المقوّم من معرفة ومتابعة كافة الخطوات التي مر بها الباحث من البداية حتى النهاية. وإن هذا الأمر من شأنه أن يسهل عملية التحضير للبحث ويساعد على تقويمه ومقارنة نتائجه مع الدراسات الأخرى.

وبصورة عامة يمكننا أن نصنف خطوات التحضير للبحث العلمي وفق الترتيب التالي:

3-1 : اختيار موضوع البحث وتحديد عنوانه :

يتم اختيار موضوع البحث من بين المواضيع التي تمس المجتمع أو تتعلق بظواهر الطبيعة وتفاعلاتها. ويجب أن يتضمن موضوع البحث مشكلة جدية تحتاج إلى حل أو علاج أو تفسير لها. وإن أهم مصادر الحصول على المواضيع البحثية التي تتضمن مشكلات جدية هي المصادر التالية:

- **المعاناة والملاحظة:** كثيراً ما يواجه الإنسان في حياته اليومية وخلال نشاطاته العملية عدداً من المشكلات والصعوبات التي تتطلب حلولاً معينة، مثل: سوء استغلال الوقت، هدر المياه، الازدحام في الشوارع، سوء المعاملة، التضخم، التلوث ...الخ.

إن مثل هذه المشكلات وغيرها قد تمر على الإنسان البسيط دون أن يعيرها أي اهتمام، لا بل يحاول أن يتكيف معها أو يسلّم بها. ولكن الشخص الباحث الحريص على تقدم وطنه ومجتمعه يتصدى لدراسة هذه المشكلات ويحاول إيجاد الحلول المناسبة لها، فيختار إحدى هذه المشكلات لتكون موضوعاً لبحثه.

- **المطالعة والمتابعة:** عندما نقوم بقراءة رواية معينة أو كتاب ما, أو نتصفح صحيفة يومية أو نراجع مجلة علمية أو أدبية، أو عندما نشاهد برنامجاً تلفزيونياً أو نسمع حواراً إذاعياً، كثيراً ما نعثر فيها على قضايا ومشكلات وتساؤلات تحتاج إلى تفسيرات وحلول . وهنا يمكن للباحث أن يولي اهتماماً خاصاً لإحدى هذه المشكلات ويختارها موضوعاً لبحثه.
- **الندوات الفكرية:** كثيراً ما تثير المناقشات خلال الندوات أو الاجتماعات أو اللقاءات قضايا ومشكلات هامة، ويطالب بعض المشاركين بتفسيرها أو إيجاد حلول ناجعة لها. ومن خلال هذه المناقشات يمكن للباحث أن يتصدى لإحدى المشكلات ويختارها موضوعاً لبحثه.

- **الدراسات السابقة:** وهي الدراسات المتخصصة التي تتناول مواضيع تقع ضمن اختصاص الباحث، حيث أن مراجعة مثل هذه الدراسات تفيد الباحث في عدة أمور هامة أهمها:

في التعرف على كل ما أنجز في مجال اختصاصه، وذلك من أجل الاستفادة من تلك الإنجازات وتوظيفها في بحثه وتفادي الوقوع في تكرار اختيار موضوع مدروس سابقاً.

كما تفيده في البحث عن تساؤلات أو مشكلات تركتها الباحثون في تلك الدراسات دون حل أو إجابة، ليختار أحداها موضوعاً لبحثه.

كما تفيده في وضع المبررات والأساليب التي دعت الباحث لاختيار موضوع بحثه وفي بيان مدى اختلافه عن المواضيع المطروحة في الدراسات السابقة.

- **وضع العنوان الدقيق:** بعد أن يختار الباحث موضوع بحثه يجب عليه أن يضع له عنواناً دقيقاً ومختصراً ومؤلفاً من طرفين مرتبطين، ويجب أن يعبر العنوان عن مشكلة البحث ويتضمن مكانها وزمانها. ويتم الاتفاق على العنوان النهائي مع الأستاذ المشرف والقسم المختص.

وأخيراً نشير إلى أنه يجب على الباحث أن يحسن اختيار الموضوع بحيث تكون مشكلته قابلة للبحث والدراسة والاختبار وأن تقع ضمن إمكانياته العلمية والمادية، فلا يختار مشكلة كبيرة أو متشعبة توقعه في متاهات معقدة وتثبط من عزيمته العلمية، ولا يتناول مشكلة لا يكون أهلاً لها، ولا يبحث في مشكلة تفوق إمكانياته المادية والشخصية.

2-3 : مراجعة وتلخيص الدراسات المرجعية السابقة:

تعتبر الدراسات السابقة من المصادر الأساسية للبحث. لذلك يجب على الباحث أن يراجع تلك الدراسات ، وأن يقوم بتلخيص محتوى كل منها بدقة بالغة، وأن يشير إلى مكان وزمان كل منها، وأن يذكر الفرضيات والنتائج التي توصل إليها الباحثون في تلك الدراسات، وأن يجري مقارنة فيما بينها، وأن يستخلص منها الجانب غير المدروس فيها، وأن يستنبط المشكلات التي تثيرها ...الخ. ويفضل التركيز على الدراسات المحلية التي تناولت موضوع البحث خلال الفترات السابقة.

وفي نهاية هذا الملخص يجب أن يذكر ويحدد مكان وزمان البحث الذي سيقوم به ومدى اختلاف مشكلته عن المشكلات المدروسة في الدراسات السابقة، وأن يقدم عرضاً يبين فيه اختلاف الفرضيات الأساسية والمناهج والأدوات المستخدمة لديه عن الدراسات المذكورة ...الخ.

وباختصار على الباحث أن يبرهن للقارئ وللجنة الحكم أن بحثه يختلف عن البحوث السابقة من حيث المشكلة والمنهج والأدوات والنتائج . وإذا لم يتحقق له ذلك فسيعد بحثه تكراراً لبحث سابق وليس فيه إضافة جديدة وبذلك لن يكون له أية قيمة علمية.

3-3 : تحديد مشكلة البحث وصياغتها بدقة:

بما أن مشكلة البحث هي المركز الذي سيدور حوله كامل البحث، لذلك يجب على الباحث أن يقوم بتحديد مشكلة بحثه تحديداً دقيقاً، بحيث يتضمن جوهر المشكلة وعناصرها المختلفة والعوامل المؤثرة عليها ومكان وزمان تواجدها.

إن الفهم الدقيق لمشكلة البحث وصياغتها بشكل واضح يساعدان الباحث على تركيز جهوده ونشاطاته على مشكلته الخاصة وتوجيهه إلى معرفة المعلومات والبيانات المتعلقة بها وارشاده إلى المصادر الأساسية لها. وكلما استطاع الباحث أن يصيغ مشكلة بحثه بعبارات لفظية دقيقة وموفقة أمكنه أن يتعامل معها ومع متغيراتها بسهولة تامة.

وهناك عدة معايير لصياغة مشكلة البحث هي:

- أن تكون الصياغة واضحة ودقيقة، ويمكن أن يتم ذلك على شكل نص كتابي أو على شكل سؤال أو أسئلة تعبر عن جوهر المشكلة وعناصرها والعوامل المؤثرة فيها . وكمثال على ذلك نأخذ: مشكلة تدني نسبة النجاح في أحد المقررات فنقول:

  إن الطلاب يعانون من تدني نسبة النجاح في مقرر كذا، وإن هذا الأمر يشكل مشكلة دراسية لهم، لأنها تنعكس على نتائج تحصيلهم العلمي وانتقالهم إلى السنوات الأعلى، وإن هذه المشكلة أصبحت ظاهرة متكررة، وفي رأينا أن هناك عدة عوامل مؤثرة فيها منها: أسلوب المدرس، اختصاص المدرس، خبرة المدرس، غياب الطلاب، سوء توقيت الامتحان، صعوبة الأسئلة، سلّم التصحيح ...الخ.

  ويمكن أن نصيغ هذه المشكلة على شكل أسئلة كما يلي:

  هل هناك تدني في نسبة النجاح فعلاً ؟

  هل يؤثر ذلك على تحصيل الطلاب ؟

  ماهي أسباب تدني نسبة النجاح في ذلك المقرر؟

  هل يعود ذلك إلى أسلوب الأستاذ ؟

  هل يعود ذلك إلى نوعية الأسئلة ؟

  هل يعود ذلك إلى غياب الطلاب ؟

  ....

  إن صياغة المشكلة على شكل أسئلة يسهل على الباحث وضع الفرضيات الاحصائية التي سيتناولها لاحقاً .

- أن يشير عند صياغة المشكلة إلى جميع العوامل والمتغيرات المتعلقة في المشكلة كالمتغير التابع والعوامل المؤثرة فيه. ففي مثالنا السابق نلاحظ أن المتغير التابع هو نسبة النجاح وأن العوامل المؤثرة فيه هي العوامل التي ذكرناها سابقاً، ولكن تلك العوامل لا تؤثر بنفس القوة على التابع، وإن

نتيجة البحث والدراسة هي التي ستحدد مقدار تأثير كل من هذه العوامل على التابع وستظهر الأسباب الكامنة وراء تدني نسبة النجاح.

– أن تتضمن الإشارة إلى الأضرار الناجمة عن تلك المشكلة وإلى حجم انتشارها في مكان تواجدها.

وأخيراً نشير إلى أن تحديد وصياغة مشكلة البحث يعتبر أمراً غير سهل على الباحثين المبتدئين، لأن الباحث المبتدئ قد لا يملك فكرة واضحة عن مشكلة بحثه وإن كل ما لديه هو فكرة عامة أو غامضة عن تلك المشكلة. لذلك يتوجب على الباحث أن يتعمق في فهم مشكلة بحثه وأن يستوعب جميع جوانبها وأن يحاول صياغتها بكل وضوح ودقة. إلا أنه يمكن للباحث عند الضرورة أن يعيد صياغة المشكلة أو أن يعدل صياغتها مع تقدم سير البحث وتبلور أفكاره، وذلك بشرط أن لا تؤدي تلك الصياغة الجديدة إلى انحراف عن العنوان أم عن الموضوع المدروس.

3-4 : تحديد أهمية البحث وأصالته:

بعد أن يقوم الباحث بتحديد مشكلة بحثه يجب عليه أن يسأل نفسه عن مدى أهمية البحث وأصالته . وبعد أن يتأكد من ذلك عليه أن يشرح ذلك في مقدمة تقريره النهائي، وأن يوضح مدى أهمية البحث وفائدته العلمية وقيمته النظرية والتطبيقية ومقدار مساهمته في تقدم المعرفة الإنسانية. كما عليه أن يشير إلى أصالة وحداثة بحثه وأن يدعم رأيه ببعض المقارنات والشواهد من الدراسات السابقة.

وحتى يكون البحث مهماً وأصيلاً يجب أن لا يدور حول مشكلة تافهة لا تستحق الدراسة والبحث (كفقدان العلكة)، وأن لا يكون تكراراً لموضوع أشبع بحثاً وتحليلاً في الدراسات السابقة (كأفكار ابن خلدون)، وأن يكون جديداً في موضوعه وعنوانه وأسلوب معالجته.

3-5 : تحديد أهداف البحث:

وهنا يجب على الباحث أن يقوم بتحديد وشرح الأهداف الأساسية والفرعية التي يتوخاها من بحثه، فيذكر الأهداف النظرية والتطبيقية للبحث ويتحدث عن النواحي التي سيهتم بها بشكل خاص وعن الجوانب التي سيغفلها مع تقديم مبررات مقنعة لأسباب ذلك الإغفال.

3-6 : تثبيت حدود مكان وزمان البحث:

وهنا يجب تحديد المكان الذي ستجري فيه دراسة المشكلة تحديداً دقيقاً وتثبيت اللحظة أو الفترة الزمنية التي ستتم فيها تلك الدراسة .

3-7 : تحديد مجتمع وعينة البحث:

يُعتبر من الضروري جداً أن يقوم الباحث بتحديد المجتمع الإحصائي الذي سيتناوله البحث وتعريفه تعريفاً دقيقاً، وبتحديد حجم العينة ونوعها وشرح كيفية سحبها ثم دراسة خواصها المميزة من جميع الجوانب للتأكد من أنها غير متحيزة وتمثل المجتمع المدروس تمثيلاً جيداً.

3-8 : تعريف المصطلحات والمتحولات:

وهنا يجب على الباحث أن يعود إلى عنوان ومشكلة البحث ويفصلها تفصيلاً دقيقاً ويشتق منها المصطلحات أو العوامل المؤثرة في الظاهرة المدروسة وأن يضع لها تسميات مناسبة، كما عليه أن يحدد المتحولات المستقلة والتابعة (قد يكون أكثر من متحول تابع) ويضع لها التسميات المعبرة عنها، ثم يقوم بتحديد واحدة القياس الملائمة لكل من المتحولات القابلة للقياس، وبتحديد سلم تدريجي لترتيب حالات كل من المتحولات النوعية غير القابلة للقياس.

3-9 : وضع فرضيات البحث وصياغتها بطريقة علمية:

بعد أن يقوم الباحث بتحديد المتحولات المستقلة والتابعة ينتقل إلى وضع الفروض (الفرضيات) التي يعتقد أنها ستجيب على التساؤلات المطروحة في المشكلة أو أنها تفسر الجوانب الغامضة فيها . وذلك بوضع فرضية مناسبة مقابل كل تساؤل في المشكلة.

وتعرف الفرضية بأنها: موقف مؤقت أو جواب مقترح لتفسير جانب غامض من المشكلة أو للإجابة على أحد التساؤلات المطروحة فيها.

ويتم صياغة الفرضيات على شكلين أساسيين هما:

- **صيغة الإثبات:** وفيها يتم صياغة الفرضية على شكل إقرار لمسألة معينة أو تأكيد على علاقة ما بين المتحولات المستقلة والتابعة، كأن نقول:

  فرضية: إن علامات الطلاب تتأثر بنسبة النجاح.

  فرضية: هناك علاقة بين عدد الحضور ونسبة النجاح.

  فرضية: إن الأسئلة كانت صعبة.

  ....

  ولكن صيغة الفروض بهذه الطريقة لا تساعد على اختبارها والتأكد من صحتها باستخدام الوسائل الإحصائية المعروفة.

- **صيغة النفي:** وهي الصيغة الملائمة للاختبارات الإحصائية التي تعتمد على ما يسمى بفرضية العدم، أي هي الصيغة التي تنص على عدم وجود فرق بين قيمتين أو على عدم وجود علاقة بين متحولين..الخ. ويتم التأكد من صحة أو خطأ تلك الفرضية بواسطة مؤشرات الاختبارات الإحصائية المعروفة. وتصاغ الفرضيات وفق صيغة النفي على الشكل التالي:

  ف$_0$: لا تتأثر محصلات الطلاب بنسبة النجاح: H_0 .

  ف$_0$: لا توجد علاقة بين عدد الحضور ونسبة النجاح: H_0 .

  ف$_0$: إن الأسئلة ليست صعبة : H_0 .

  ....

ويرمز لفرضية العدم بالرمز ف0 (H_0) وتقابلها فرضية معاكسة تسمى الفرضية البديلة ويرمز لها بالرمز ف1 (H_1) . وسنتعرض إلى هذه الأمور لاحقا عند دراستنا لاختبارات الفرضيات.

وهنا يجب أن نشير إلى أن أهداف البحث وطبيعة المشكلة ونوعية متحولاتها تلعب دوراً كبيراً في تحديد نوع الفرضيات وعددها، وبالتالي في تحديد المؤشرات الإحصائية لاختبارها. وفي كثير من الحالات يمكننا أن نضع فرضية عامة للبحث، ثم نفرعها إلى فرضيات جزئية لتسهيل عمليات الاختبارات الإحصائية.

3-10 : تحديد المنهج والأساليب:

وهنا يقوم الباحث بتحديد المنهج العلمي المتبع في إنجاز البحث ويشير إلى الأساليب المستخدمة في التحليل والاستنتاج. ويمكن للباحث أن يتبع أكثر من منهج في الإنجاز وأن يستخدم أكثر من أسلوب في التحليل والاستنتاج. فمثلاً يمكن أن يتبع الباحث المنهج الوصفي التحليلي ومنهج المسح الإحصائي في دراسة نتائج الطلاب وأن يستخدم عدداً من الأساليب الإحصائية، كحساب النسب المئوية والمعدلات ودراسة الارتباط وتطبيق الاختبارات واستخدام النماذج القياسية والبرامج الحاسوبية ...الخ.

3-11 : صياغة مخطط البحث:

إن مخطط البحث يجب أن يتوافق مع عنوان البحث ولا يشكل خروجاً عنه، وبشكل عام يمكن أن يكون على الشكل التالي:

- مقدمة: وتتضمن جميع الأمور المنهجية السابقة وتلخيصاً للدراسات المرجعية .
- فصل أول: ويتضمن عرضاً نظرياً لإطار المشكلة.
- فصل ثاني: ويتضمن قضايا جمع البيانات اللازمة.
- فصل ثالث: ويتضمن معالجة وتحليل البيانات واستخلاص النتائج.
- الاستنتاجات والمقترحات.
- الفهرس: ويتضمن ترتيب الفصول والفقرات في الرسالة أو الأطروحة (ويوضع في بداية االرسالة).

ويفضل أن يتم تقسيم كل فصل إلى فقرات معينة تعالج أمور محددة في البحث. وفي الحقيقة لا يوجد مخطط موحد صالح لجميع البحوث وإن ضرورات البحث هي التي تفرض عدد الفصول وتحدد عدد الفقرات في كل فصل. وهنا ننصح أن لا يتوسع الباحث في عدد الفصول ولا في عدد الفقرات الداخلية إلا حسب ما تقتضيه الضرورة.

3-12 : إعداد قائمة المراجع والمصادر:

يجب أن تكون هذه القائمة مكتوبة حسب الأصول المكتبية المعتمدة في الجامعة أو المؤسسة ومرتبة حسب الترتيب الهجائي للمؤلفين أو المشرفين مع مراعاة أن تكون هذه المراجع حديثة ومن أقطار مختلفة وبلغات متعددة. ويجب التركيز على المراجع المحلية حسب توفرها.

3-13 : كتابة البحث وتقديمه للمناقشة والاعتماد:

ويتم ذلك بتقديم مشروع رسالته أو إطروحته المتضمن جميع الأمور السابقة إلى القسم المختص لمناقشته وإجراء ما يراه ضرورياً من تعديلات عليه ثم إعادة مناقشته واعتماده وتشكيل لجنة الحكم لإقراره ومنح الطالب الشهادة المسجل عليها .

الفصل الرابع
خطوات وطرائق جمـع البيـانـات

إن مرحلة جمع البيانات تعد من أطول وأصعب مراحل البحث العلمي، فهي تحتاج إلى تنظيم كثير من الأمور ومتابعة تنفيذها، وإجراء القياسات أو الحصول على المعلومات بطرائق مختلفة ثم تسجيلها في جداول مناسبة. لذلك فإننا سنستعرض هذه الأمور من خلال الفقرتين التاليتين :

4-1: خطوات جمع البيانات:

1- **تحديد طريقة الجمع:** وهناك عدة طرائق لجمع البيانات هي: القياس المباشر، الاستخلاص، الملاحظة، المقابلة، الاستبيان. وبتحديد طريقة جمع البيانات تتحدد جميع الأمور الأخرى المرتبطة بها.

2- **وضع خطة الجمع:** ويتم ذلك بتحديد مراحل عملية الجمع وفتراتها الزمنية ولوازمها المادية وتكاليفها المالية وعناصرها البشرية.

3- **تصميم الاستمارة اللازمة:** وهي عبارة عن جملة من الأسئلة المنسقة أو الجداول المنظمة والتي تحتاج إلى تعبئة من قبل العاملين عليها أو المعنيين بها. ويجب أن تكون الاستمارة ملائمة لطريقة جمع البيانات، فالاستمارة المستخدمة في طريقتي القياس والاستخلاص تكون مؤلفة من جداول منظمة وتشمل جميع المؤشرات المطلوبة، أما الاستمارة المخصصة لطريقة الملاحظة فتتضمن حالات أو قيم افتراضية متعددة ويقوم الباحث بوضع إشارة على أحدها عندما يلاحظ ظهور تلك الحالة أو تحقق تلك القيمة، ولكن الاستمارة المناسبة لطريقة المقابلة فيجب أن تتضمن أسئلة مباشرة ومحددة وسهلة الفهم والجواب، غير أن الاستمارة الملائمة لطريقة الاستبيان يجب أن تتضمن عدا عن الأسئلة المباشرة شروحات بسيطة لبعض المفاهيم والمصطلحات وتعليمات معينة لكيفية تعبئتها وإعادتها من قبل الشخص المبحوث.

وبصورة عامة يجب على الباحث عند تصميم الاستمارة الإحصائية مراعاة الأمور التالية:

- مراجعة وحصر المتحولات والبيانات اللازمة لبحثه ثم تجميعها ضمن مجموعات متجانسة، مثل: مجموعة المعلومات العامة، مجموعة المؤشرات الاقتصادية، مجموعة المؤشرات الصحية، مجموعة المؤشرات النفسية ...الخ.
- تقسيم الاستمارة إلى عدة أقسام منفصلة بحيث يتناول كل قسم مجموعة من المجموعات السابقة أو جانباً من البيانات المطلوبة.

- صياغة الأسئلة بشكل واضح ومختصر وبسيط وتجنب وضع الأسئلة المبهمة أوالطويلة أوالموحية بأي جواب معين.
- الحرص على أن تكون الأسئلة متناسبة مع مستوى المبحوث وتمس قضاياه الأساسية وتخدم هدف البحث، وأن لا تتضمن عبارات محرجة أو جارحة لأي طرف.

وفي كثير من الأحيان يتم تصميم الاستمارة الإحصائية بحيث تتضمن عدة خيارات أو بدائل للجواب على معظم الأسئلة . لذلك يميز الدارسون نوعين من الأسئلة هما:

- الأسئلة ذات الأجوبة المغلقة أو المحددة.
- الأسئلة ذات الأجوبة المفتوحة أو الحرة.

فالأسئلة ذات الأجوبة المغلقة تشترط أن يكون لكل سؤال عدد محدد من خيارات الأجوبة، وما على المبحوث إلا أن يقوم بوضع إشارة معينة على الخيار الذي يراه مناسباً لحالته.

وتبدأ هذه الخيارات بوضع جوابين فقط على شكل:

نعم	لا

أو عل شكل آخر كما يلي:

موفق	غير موافق

وكمثال على ذلك نأخذ السؤالين التاليين:

هل قمت بزيارة مدينة تدمر؟

نعم	لا

ما رأيك أن تذهب معي؟

موافق	غير موافق

....

ويمكن أن يكون عدد الخيارات ثلاثة أجوبة أو أكثر على شكل (كتابي أو رقمي) كما يلي:

ما رأيك في مشاركة المرأة في العمل؟	موافق	حيادي	غير موافق		
ما رأيك في إنشاء المصنع؟	موافق جداً	موافق	معارض	معارض جداً	
ما رأيك في البرنامج المفتوح ؟	جيد جداً	جيد	وسط	ضعيف	سخيف

ما هي درجة حبك للرياضة ؟	5	4	3	2	1	
ما هي درجة احترامك للسياسة؟	5	4	3	2	1	0

وهنا نشير إلى أن تعدد الخيارات قد يؤدي إلى إرباك المبحوث والتأثير على إجابته، فقد يتحيز المبحوث بشكل عفوي إلى أحد الخيارات: كأن يختار الجواب الأول أو الجواب الألطف أو الجواب الأسوأ، وتحدث مثل هذه الأمور عندما يكون رأي الباحث غير ناضج حول موضوع السؤال، لذلك ننصح أن يكون شكل الخيار رقمياً, لأنه لا يوحي للمبحوث بأي جواب ولأن المسافات بينها متساوية، بينما المسافات في الخيارات الكتابية غير معروفة وغير متساوية، فالمسافة بين موافق جداً وموافق لا تساوي المسافة بين الموافق والمعارض ...الخ, وهذا يشكل خللاً في هيكل السؤال.

كما يفضل أن يكون عدد الخيارات قليلاً وزوجيا ويكتفي بخيارين أو أربعة خيارات على الأكثر، لأن ذلك يضمن عدم تحيز الباحث نحو الخيار الأوسط، كما ننصح عدم استخدام الكلمات التي قد تسيء لشخص المبحوث مثل: غير مهتم أو غيرذلك.

أما الأسئلة ذات الأجوبة المفتوحة فتصاغ , بحيث يتم ترك الخيارات مفتوحة أمام المبحوثين, وكمثال عن ذلك نأخذ الأسئلة التالية:

ما هي المشكلات التي تعاني منها في الدراسة ؟

ما هي المقررات التي تميل إليها ؟

من هم أفضل الأساتذة برأيك ؟

من هم أحسن المرشحين لديك ؟

ويقوم المبحوثون بالإجابة على هذه الأسئلة بشكل حر وبدون أي توجيه أو تدخل من قبل الباحث .

4- **تدريب الكوادر وتأمين اللوازم** : يفضل أن يقوم الباحث بجمع البيانات بنفسه, ولكن إذا اقتضى الأمر الاستعانة ببعض الكوادر المختصة بجمع البيانات, فإنه يجب تدريبهم على جمع البيانات وفق الطريقة المحددة لذلك،لإن هذا التدريب يضمن للباحث الحصول على معلومات دقيقة ويمكن الكوادر من معالجة المشكلات التي تعترضهم عند الجمع. وهنا يجب على الباحث تأمين جميع اللوازم المادية والمالية للقيام بعملية الجمع بدون أية عوائق.

5- **إجراء اختبار تجريبي**: وذلك للتأكد من صلاحية الاستمارة ولتدريب الكوادر المكلفة على استخدامها ولتدعيم الثقة فيهم وفي صحة المعلومات التي سيحصلون عليها.

6- **إجراء التعديلات اللازمة على الاستمارة** وتدارك الأخطاء والنواقص التي أظهرها الاختبار التجريبي ثم الانطلاق في عملية الجمع النهائي.

4-2: طرائق جمع البيانات: إن طرائق جمع البيانات هي:

4-2-1 : **طريقة القياس المباشر**: وفيه يقوم الباحث أو المكلفون بجمع البيانات بإجراء القياسات بشكل مباشر للمتحولات المؤثرة في الظاهرة المدروسة وللمتحولات التابعة لها، ويتم تسجيل هذه القياسات في جداول مصممة خصيصاً لذلك، وفي هذه الحالة يجب تحديد واحدة القياس بشكل مناسب لكل متحول على حدة، أما إذا كان المحول نوعياً غير قابل للقياس فيجب وضع سلم تدريجي لترتيب حالاته المختلفة.

4-2-2 : **طريقة الاستخلاص المباشر**: وفيه يقوم الباحث أو المكلفون بجمع البيانات باستخلاص القياسات أو المعلومات من السجلات المعتمدة والموثوقة ونقلها إلى استمارات أو جداول مصممة خصيصاً لذلك، وهنا يجب مراعاة جميع التغيرات المكانية والزمانية التي يمكن أن تكون قد حدثت خلال فترة الدراسة وإجراء التعديلات اللازمة لتصحيح المعلومات وتنسيقها.

3-2-4 : **طريقة الملاحظة:** تعد الملاحظة من الطرائق المعتمدة في عمليات البحث العلمي . فمنذ القدم لاحظ الإنسان ظهور شرارة من تصادم الأحجار الصلبة فقادته هذه الملاحظة إلى اكتشاف النار، كما لاحظ أن الأخشاب تطفو على سطح الماء فتوصل إلى اختراع القارب ثم السفينة، ولاحظ اهتزاز الغطاء عند غليان الماء فاكتشف القوة البخارية واخترع القطارات، وعندما لاحظ نيوتن سقوط التفاحة من الشجرة اهتدى إلى قانون الجاذبية الأرضية ...الخ.

ولقد لعبت الملاحظة دوراً مشابهاً في ميادين العلوم الاجتماعية، فكان الإنسان يلاحظ سلوك وتصرفات أخيه الإنسان ويستنتج ما هو حسن وما هو سيئ، وما هو ضار وما هو نافع.

وبذلك ساهمت الملاحظة في ترسيخ وتطوير النظم الأخلاقية لدى الشعوب المختلفة.

ونظراً لأهمية الملاحظة كطريقة للبحث العلمي قام المختصون بتنظيم عملياتها وبوضع أسس لها لتؤدي دورها بشكل فعال في تطوير البحث العلمي وميّزوا بين عدة أنواع للملاحظة العلمية هي:

- الملاحظة البسيطة: وهي التي يقوم فيها الباحث بملاحظة وتسجيل مجريات الظاهرة المدروسة كما هي على الواقع وبدون أي إعداد مسبق أو تدخل في العوامل المؤثرة فيها.
- الملاحظة المنظمة: وهي التي يقوم فيها الباحث باتباع مخطط مسبق لمراقبة الظاهرة، ويمكن له أن يتحكم ببعض العوامل المؤثرة فيها وأن يستخدم مختلف الأجهزة والمعدات التي تساعده في عمليات المراقبة والتسجيل، وأفضل مثال على ذلك عمليات الرصد الفلكية، أو مراقبة الأسعار في السوق ...الخ.
- الملاحظة المباشرة: وهي التي يقوم الباحث فيها بالدخول إلى صفوف الجماعة المدروسة (بشكل علني أو سري) ومشاركتهم مختلف نشاطاتهم الحياتية، فهو بذلك يلعب دورين: الأول دور عضو في الجماعة والثاني دور باحث ودارس لأحوالها. إن هذا الأسلوب يمكّن الباحث من الحصول على بيانات صحيحة ومباشرة ويستخدم في دراسة مشكلات اجتماعية عديدة: كأحوال المساجين وحياة القبائل ونشاطات العصابات ...الخ.
- الملاحظة غير المباشرة: وهي التي يلعب فيها الباحث دور المتفرج ويقوم بالمراقبة بواسطة المشاهدة والاستماع . ويسجل مختلف المواقف دون المشاركة الفعلية فيها.

وتتصف الملاحظة العلمية بالمزايا التالية:

- أنها أسهل طريقة لجمع البيانات ولها استخدامات في مجالات كثيرة.
- أنها تسمح بجمع معلومات مباشرة وبظروف طبيعية مألوفة.
- أنها لا تتطلب جهداً كبيراً بالمقارنة مع الطرائق الأخرى.
- أنها تعتمد على الاستنتاجات الحسية أكثر من الاستنتاجات النظرية.

ويسجل على الملاحظة عدة عيوب هي:

- قد يتعمد المبحوثون إعطاء انطباع جيد عن سلوكهم وتقاليدهم وخاصة عندما يشعرون بأن الباحث يقوم بمراقبتهم وتسجيل كافة حركاتهم.
- قد يتطلب تطور بعض الأحداث زمناً طويلاً، مما يكلف الباحث وقتاً مماثلاً لمراقبتها حتى يستطيع تسجيل الملاحظات عنها.
- قد تكون التفسيرات التي يضعها الباحث حول تغيرات الظاهرة متحيزة ومتأثرة بأهوائه الشخصية وبثقافته المعرفية.

4-2-4 : **طريقة المقابلة:** المقابلة هي لقاء ومحادثة بين الباحث والمبحوث يهدف إلى جمع البيانات اللازمة عن الظاهرة المدروسة. وقبل إجراء المقابلة يجب الإعداد لها إعداداً دقيقاً بحيث يتم تأمين الأمور التالية:

- تحديد أهداف المقابلة.
- تحديد أسئلة المقابلة ضمن استمارة خاصة بذلك.
- تحديد الأفراد الذين ستشملهم المقابلة.
- التدريب على إجراء المقابلة.

وحتى تكون المقابلة فعالة وناجحة يجب أن يتحقق الانسجام بين أطرافها الثلاثة: الباحث والمبحوث واستمارة البحث، وهذا يتطلب من الباحث أن يراعي الأمور التالية:

- أن يلتزم بالمواعيد المحددة مع المبحوث، وأن يظهر بمظهر لائق أمامه.
- أن يعامل المبحوث بكل ودٍ واحترام، وأن يؤكد له أن هدف البحث علمي صرف وإن المعلومات التي سيدلي بها ستبقى سرية تماماً.
- أن يفتتح الحوار بأسئلة عامة ومشوقة وذلك بهدف إزالة تحفظ أو تخوف المبحوث وتشجيعه على الكلام.
- أن يهتم بجميع ما يقوله المبحوث وعدم مقاطعته أثناء الحديث أو الانشغال بأشياء أخرى.
- أن يقوم بنفسه أو بمساعدة أحد الكوادر المدربة بتسجيل محتوى الحديث أثناء المقابلة وليس بعدها، ويجب الاستئذان من المبحوث عند استخدام أجهزة التسجيل.

وحتى يتحقق الانسجام الكامل خلال المقابلة يفترض أن تتوفر في المبحوث الشروط التالية:

- الاستعداد لإجراء المقابلة دون أي ضغط أو تخوف.
- إدراك هدف البحث ودوره في خدمة المصلحة العامة.
- المرونة في تقبل أسئلة الباحث وتعليماته والتجاوب معها.
- القدرة المعرفية الكافية للإجابة على الأسئلة المطروحة عليه بكل صدق وموضوعية.

أما الشروط التي يجب أن تتوفر في الاستمارة فلقد ذكرناها سابقاً وأهمها:

- أن تكون الأسئلة واضحة وقصيرة وغير موحية بأي جواب ومناسبة للمستوى التعليمي للمبحوثين ولا تتطلب جهداً عقلياً للإجابة عليها.

وتتصف طريقة المقابلة بعدة مزايا أهمها:

- إنها تصلح لجمع البيانات في الأوساط الاجتماعية غير المتعلمة في الريف والمدينة.
- إنها تساعد على فهم الأسئلة وشرح ما هو غامض فيها دون حذف أو تحوير.
- إنها تضمن أخذ رأي المبحوث مباشرة والحصول على معلومات دقيقة منه.
- إنها تضمن الإجابة على جميع الأسئلة الواردة في الاستمارة.
- إنها تسمح للباحث بإمكانية تقديم أو تأخير الأسئلة لتسهيل عمليتي الفهم والإجابة.

أما عيوب طريقة المقابلة فنلخصها بما يلي:

- إنها تتطلب كثيراً من الجهود الشخصية والتكاليف المادية والمالية.
- إنها تحتاج إلى عدد كبير من جامعي البيانات ويلزمها وقت طويل لإجراء الجمع.
- إن اللقاء المباشر مع المبحوث قد يبعث لديه بعض الإحراج فيدلي بإجابات غير موضوعية أو متحيزة، أو يتهرب من الإجابة على بعض الأسئلة الحساسة.

5-2-4 : **طريقة الاستبيان:** تقوم طريقة الاستبيان على أسلوب جمع البيانات بالمراسلة، وتعتمد على إعداد استمارة خاصة تضم الأسئلة التي تخدم هدف البحث، وترسل إلى المبحوثين, الذين تم سحبهم في عينة البحث بواسطة استخدام إحدى الوسائل التالية:

- البريد العادي: وهو وسيلة مناسبة وغير مكلفة، ويستخدم البريد عندما يكون أفراد العينة موزعين في أماكن متعددة ومتباعدة . ويستحسن إرسال مغلف الإعادة جاهزاً مع الاستبيان.
- التسليم باليد: وهي تصلح عندما يكون إفراد العينة متواجدين في مكان واحد، كطلاب الجامعة أو عمال المؤسسة أو تلاميذ المدرسة. وهنا يقوم الباحث بتسليم الاستمارة إلى أفراد العينة ثم يعود إليهم بعد فترة معينة ليأخذها منهم .
- النشر والإعلان: وفيها يتم نشر الاستبيان في الصحف أو المجلات، أو إعلانه في الراديو أو التلفزيون أو نشره على وسائل التواصل الاجتماعي، والطلب من القراء أو المستمعين أو المشاهدين أو المتابعين أن يبادروا إلى الإجابة على الأسئلة المطروحة وإرسالها بريدياً أو إبلاغها هاتفياً أو الكترونياً وهنا يجب أن تكون الأسئلة قليلة ومغلقة.
- البريد الالكتروني: وبواسطته يمكن أن يتم توجيه الأسئلة عبر شبكة الانترنيت إلى عينة من المشتركين فيها، والطلب منهم الإجابة عليها وإرسالها إلى المصدر الأصلي.

إن أهم مزايا طريقة الاستبيان هي:

- إمكانية إرسال الاستمارة إلى عدد كبير من المبحوثين دون تحمل نفقات كبيرة.
- إنها تسمح بإعطاء المبحوث فرصة للتفكير قبل الإجابة على الأسئلة المطروحة.
- إنها تستبعد تدخل الباحث وإحراج المبحوث عند الإجابة.
- إن تسليم الاستمارة باليد يحث المبحوثين على التجاوب معها ويحفزهم على الإجابة.

أما عيوب طريقة الاستبيان فنلخصها بما يلي:

- إنها لا تستخدم إلا مع الأفراد المتعلمين.
- إن بعض الأجوبة قد تكون غير دقيقة نتيجة لعدم وضوح الأسئلة أو لتسرع المبحوث.
- إن عدداً كبيراً من المبحوثين لا يجيب على بعض الأسئلة لعدم فهمها أو لعدم اهتمامهم بها.
- إن عدداً كبيراً من المبحوثين لا يعيد الاستمارة ولا يهتم بها أصلاً.

الفصل الخامس

قواعـد كتابـة الرسالة أو الأطـروحـة

يبدأ الباحث في التفكير بكتابة تقريره البحثي أو رسالته أو أطروحته منذ اللحظة الأولى لشروعه في إجراء البحث، ويبقى هذا الأمر يشغل ذهنه طيلة فترة عمله فيه. ولكن بما أن ذلك التقرير يتألف عادة من مقدمة وإطار نظري وتطبيق تجريبي وخاتمة وبعض الملاحق، فإن الباحث يمكنه أن يستفيد من وقته الموازي لسير العمليات التجريبية، فيقوم بتوفير المصادر النظرية ودراسة المراجع الأساسية ومراجعة الدراسات السابقة. وخلال هذه المراجعات عليه أن يضع إشارات خاصة وأن يكتب ملاحظات موجزة عن كل ما يهمه في كتابة التقرير. وبعد أن يستوعب كامل الجوانب النظرية المتعلقة بموضوع بحثه يجب عليه أن يشرع بكتابة الإطار النظري لبحثه، ويمكنه أن يستفيد من كل ما كتبه حول البحث في مرحلة التحضير ليضعه في مقدمة التقرير أو في مكان مناسب آخر. وعند الكتابة عليه أن يراعي بنود المخطط المعتمد لبحثه وأن لا يخرج عنه إلا إذا اقتضت الضرورة وبعد أخذ موافقة المشرف على ذلك.

وعادة تتم كتابة االرسالة أو الأطروحة على شكل فصول متتالية ولها أسماء محددة، ويتفرع كل منها إلى فقرات وبنود ذات عناوين مختصرة. ويمكن للباحث أن ينظم هذه الفصول حسب حاجته وخبرته واهتماماته العلمية .

5-1: الهيكل العام للبحث: ويجب أن يتألف من العناصر التالية:

5-1-1:المقدمة :

يجب أن تتضمن تمهيداً للبحث وتلخيصاً للدراسات السابقة وتحديداً لمشكلة البحث و أهدافه وأصالته وعلاقته بمجالات المعرفة المرتبطة فيه، وأن تشتمل على تعريف للمجتمع والعينة وتحديد للمكان والزمان وشرح للمنهج والأساليب المستخدمة في المعالجة. كما يمكن للباحث أن يضع في تلك المقدمة وصفاً لبعض المراحل التي مرّ بها البحث، أو أن يشرح بعض الظروف المحيطة به، أو أن يتعرض إلى غيرها من الأمور التي يرى ضرورة لذكرها في المقدمة. وهنا نشير إلى أنه يجب على الباحث أن يعيد النظر في ما كتبه في المقدمة في إطار معاناته العملية وعلى ضوء النتائج التي توصل إليها.

وإن هذا الأمر جعل كثيراً من الباحثين يؤجل كتابة المقدمة العامة ويجمد التفكير فيها إلى ما بعد الانتهاء من جميع عمليات البحث.

5-1-2: الإطار النظري:

وفيه يستعرض الباحث مختلف النظريات والآراء المتعلقة بموضوع بحثه ويحدد موقفه منها، ويشير إلى ميزات ومناهج ونتائج ونواقص كل منها، ويحدد اختلاف منهج بحثه عنها. ثم ينتقل إلى عرض

مشكلة بحثه، فيعرف متحولاتها، ويصيغ فرضياتها، ويضع تساؤلاتها. وبعدها يقوم بتحديد طرائق جمع البيانات وأساليب معالجتها ومؤشرات اختبار فرضياتها وكيفية دراسة العلاقات بين متحولاتها وتحليل النتائج المتوقعة منها . ويمكن للباحث أن يفصّل هذه الأمور أو يختصرها وأن يخصص لها فصلاً أو أكثر.

5-1-3: الإطار التطبيقي:

وفيه يقوم الباحث بعرض ومعالجة وتحليل البيانات التي تم جمعها . فيدرس صفات وميزات العينة المسحوبة ليتأكد من عدم تحيزها من جميع الجوانب. ثم يقوم بمعالجة تلك البيانات وحساب المؤشرات التي تخدم هدف البحث: كالمتوسطات والتباينات والنسب والمعدلات وتحليلها وإجراء عمليات اختبار الفرضيات واتخاذ القرارات المناسبة حولها ودراسة الارتباط والانحدار بين المتحولات والتنبؤ بتغيراتها ...الخ.

ومن خلال المعالجات السابقة يتوصل الباحث بعد كل عملية إلى نتيجة محددة، وعليه أن يقوم بتلخيص تلك النتائج وتفسيرها تفسيراً علمياً دون أية مبالغة أو تحيز، وأن يحرص أن لا تخرج تعميماته عن حدود المجتمع المعرّف في البحث.

5-1-4: الاستنتاجات والمقترحات:

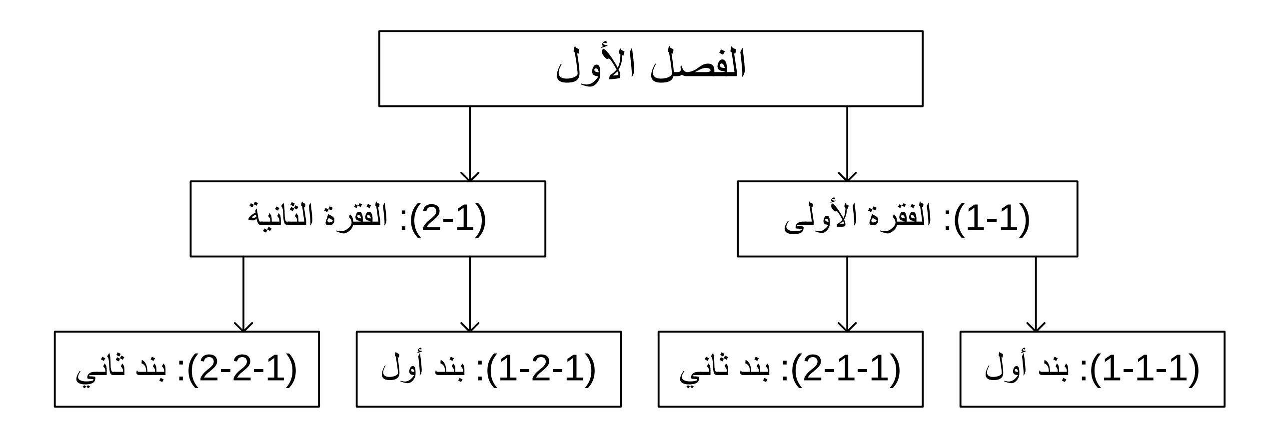
بعد أن يقوم الباحث بتلخيص النتائج، التي توصل إليها يقوم باستخلاص الاستنتاجات منها، ويعرضها حسب أهميتها أو تسلسلها، ثم يتقدم ببعض المقترحات والتوصيات التي يراها مناسبة ومنسجمة مع النتائج السابقة. ويشترط في المقترحات أن تكون محددة تحديداً جيداً ومصاغة بطريقة مختصرة وعلى شكل بنود منفصلة وتعالج قضايا معينة ومنبثقة عن النتائج التي أفرزها البحث.

5-2: قواعد الكتابة والطباعة:

إن الاهتمام منذ البداية بتنظيم عمليات الكتابة والطباعة تساعد على إخراج التقرير بشكل منسق وجميل، وتعمل على توضيح هيكله وإبراز مضمونه وتفهم أفكاره. في الحقيقة لا يكفي أن يبذل الباحث جهداً كبيراً في عمليات البحث والتحليل فقط، بل من المهم جداً أن يقوم بعرض هذا الجهد بطريقة منظمة تؤمن له إيصال مضمون بحثه إلى الآخرين. ومن الأمور التي يجب أن يهتم بها الباحث عند الكتابة والطباعة الأمور التالية :

1- أن تتم الكتابة بلغة سليمة وبأسلوب علمي سهل وبعيد عن المبالغة والتحيز، وأن تبقى الأفكار محصورة ضمن إطار البحث.

2- أن يتم وضع عناوين مناسبة للفصول والفقرات والبنود وطباعتها بشكل بارز بعد إعطائها أرقاماً مناسبة ومتفرعة بشكل عنقودي وموضوعة ضمن قوسين أو بلا قوسين، بحيث يشير الرقم الأول إلى الفصل، والثاني إلى الفقرة، والثالث إلى البند، وذلك وفق الشكل التالي:



ويمكن للباحث أن يستخدم في الترقيم الأحرف الأبجدية أو الأرقام العددية ضمن البنود الفرعية . وهنا نذكر الباحثين بأن لا يقل عدد الفقرات في كل فصل عن فقرتين وأن لا يقل عدد البنود في كل فقرة عن بندين.

3- أن يتم الالتزام التام بالدقة والأمانة العلميتين، بأن يشير الباحث إلى جميع الأفكار أو الآراء التي يستعيرها من الآخرين لتدعيم رأيه أو يستفيد منها لتوضيح قضية في بحثه . وتتم الإشارة إلى تلك الأفكار المستعارة بعدة أساليب هي:

- وضع رقم المرجع ضمن قوسين متوسطين خلال النص كالتالي: [14]، أو وضع اسم المؤلف وعام النشر ضمن هذين القوسين كالتالي: [حكيم، 1985]. ويستخدم هذا الأسلوب عندما تكون الاستفادة عامة ويقصد منها إرشاد القارئ إلى المرجع الذي يعالج قضية معينة . وهنا يشترط أن يكون المرجع وارداً في قائمة المراجع وتحت نفس الرقم.
- الإشارة خلال النص إلى المرجع الذي تم الاقتباس منه ووضعه ضمن قوسين متوسطين مع ذكر رقم الصفحة التي أخذت منها الفكرة المقتبسة كما يلي: [هيكل، 1985- ص. 125]. وهنا يجب وضع هذه الإشارة بعد الفكرة المقتبسة مباشرة.
- وضع الفكرة المقتبسة ضمن قوسين صغيرين أو هلالين يتبعهما إشارة أو رقم يدل على المرجع المذكور في نهاية الصفحة كالتالي: " إن الإدارة فن وعلم ..." 1، (إن النموذج القياسي هو تعبير رياضي عن العلاقة بين متحولين أو أكثر) 2 . وهنا يجب أن يذكر المرجع المقصود في أسفل الصفحة وتحت خطٍ مسطّر وفق الأصول المتبعة التي سنذكرها لا حقاً.

4- أن يتم الاهتمام بترتيب العلاقات الرياضية وبتنظيم الجداول الاحصائية ووضع أرقامها المزدوجة وعناوينها الدقيقة فوقها وتثبيت مصادرها تحتها، وإعداد الرسوم البيانية والأشكال التوضيحية ووضع عناوين مناسبة لمضامينها وإعطائها أرقاماً متسلسلة ومزدوجة، بحيث يشير الرقم الأول إلى الفصل، والثاني إلى رقم العلاقة أو الجدول أو الشكل كما يلي: جدول (2-3): التركيب العمري للسكان في سوريا عام 1994. أو على النحو: شكل (3-4): المخطط العامل لتنفيذ البناء .

وهنا نؤكد على أن رقم الجدول وعنوانه يجب أن يكتبا في أعلى الجدول أما مصدره فيثبت تحته مباشرة ويكتب حسب الأصول . أما رقم الشكل وعنوانه فيوضع تحته مباشرة.

5- أن يتم ترقيم الصفحات بأرقام متسلسلة وأن تتم طباعتها في وسط وأسفل (أوأعلى) الصفحة، ويجب أن يبدأ الترقيم من المقدمة حتى نهاية التقرير.

وهنا يجب الانتباه عند تنسيق الطباعة إلى شرط أساسي هو: أن تبدأ الفصول بصفحة جديدة وبأرقام فردية لتقع على يسار القارئ العربي (وعلى يمين القارئ الأجنبي).

6- أن يتم إعداد قائمة منظمة بالمراجع المستخدمة: وهنا يجب على الباحث أن يقوم أولاً بتصنيف تلك المراجع حسب أنواعها (عربية، أجنبية، بحوث، منشورات)، ثم يقوم بترتيبها حسب الأحرف الأبجدية لكنية المؤلف (بعد تجاهل ال التعريف إن وجدت) وكتابتها في القائمة حسب الأصول التالية:

- **المراجع العربية**: وتكتب على النحو التالي:

رقم التسلسل- كنية المؤلف،اسماه الأول والثاني . عام النشر- عنوان المرجع بخط مائل أو مسطر. الطبعة، دار النشر، المدينة، البلد.

وكمثال على ذلك نأخذ المرجع التالي:

51- درويش، عدنان أحمد . 1979- ديوان الجواهري . الطبعة الأولى، وزارة الثقافة والإرشاد القومي، دمشق، سورية.

- **المراجع الأجنبية**: وتكتب كما يلي:

رقم التسلسل- كنية المؤلف بالأحرف الكبيرة، الحرف الأول من اسميه وبعد كل منهما نقطة . سنة النشر- عنوان الكتاب بحرف مائل أو مسطر ، الطبعة، دار النشر، بلد النشر. وكمثال على ذلك نأخذ المرجع التالي:

21- SIMONDS, N.M. 1979- *Evolution of Corp Plants*. 2ed.Ed. Longman Group Limited, London.

- **البحوث المنشورة في مجلات باللغة العربية والأجنبية**: وتكتب كما يلي:

رقم التسلسل- الكنية، إسماه الأول والثاني. عام النشر- عنوان البحث بحرف مائل أو مسطر. اسم المجلة، بلد النشر، رقم المجلد، رقم العدد، ويتبعه أرقام الصفحات الخاصة بالبحث ضمن المجلة.

وكمثال على ذلك نأخذ البحثين التاليين:

31- جواهري، محمد زهير . 1985- *التدابير العلاجية في الآلام الرأسية*، المجلة الطبية العربية، نقابة الأطباء، سورية، العدد 86 ص. 54-59 .

4- MAHOWALO, M. 1982- *The Image of J in the EHP Sequence*. Annals of Mathematics, U.S.A., Vol. 116, No. 1, pp. 65-112 .

- **المنشورات التي ليس لها مؤلف محدد:** وترتب حسب الأحرف الأبجدية لعناوينها (بعد تجاهل آل التعريف إن وجدت) كما يلي:

الرقم- عنوان المنشور . عام النشر، مؤسسة النشر، بلد النشر.

وكمثال على ذلك نأخذ المنشورات التالية:

12- المجموعة الإحصائية السورية . 1998، المكتب المركزي للإحصاء، سورية .

-**المنشورات الالكترونية:** وترتب حسب نوعها وفق التصنيفات السابقة مع ذكر الموقع الالكتروني التي أخذت منه وتسجيل تاريخ الدخول.

....

7- أن يتم إعداد الملاحق الضرورية، كالاستمارات الإحصائية والجداول التفصيلية والوثائق التاريخية، والتي قد تتضمن بعض المعلومات الهامة بالنسبة للبحث. ويتم ترتيب هذه الملاحق حسب أهميتها ثم إعطائها أرقاماً متسلسلة للإشارة إليها في متن البحث.

8- أن يتم إعداد فهرس للمحتويات، وفيه يقوم الباحث بعرض عناوين الفصول والفقرات المتعلقة بها وإعطائها نفس الأرقام التي وردت في متن التقرير. وهنا يجب على الباحث أن يحرص على طباعة عناوين الفصول بخط متوسط وبارز إلى اليمين، بينما يجب طباعة العناوين الفرعية بخط عادي ومنحسر إلى اليسار. كما عليه أن يضع أمام عنوان كل فصل أو فقرة رقم صفحته في متن التقرير.

9- أن تتم طباعة الغلاف بشكل فني وأنيق، بحيث يتم إظهار اسم الجامعة والكلية والقسم في أعلى يمين الصفحة. ثم يكتب العنوان بخط عريض في الوسط الأعلى للصفحة، مع الإشارة تحته وبخط صغير إلى أن هذا البحث مقدم لنيل شهادة كذا في اختصاص كذا. ويوضع الاسم الثلاثي للباحث تحت العنوان بقليل، ثم يوضع اسم المشرف (أو المشرفين) في الوسط الأدنى من الصفحة، وفي أسفل الصفحة يثبت عام الطباعة ومكانه.

10- أن تخصص صفحة بيضاء بعد الغلاف لعرض قرار تشكيل لجنة الحكم وأسماء أعضائها وتاريخ المناقشة. ويمكن للباحث أن يخصص صفحة أخرى للشكر وثالثة للإهداء.

11- أن يتم تجليد التقرير تجليداً فنياً للحفاظ عليه من التلف ولتسهيل حفظه في المكتبات، وهنا يجب إعادة طباعة محتويات الغلاف على الجلد الخارجي وترك صفحة بيضاء بين الجلد والغلاف.

الملحق (1)

تعليمات وإرشادات الطباعة والإخراج المعتمدة في جامعة تشرين

وهي متوافقة مع دليل التوثيق المعتمد في جمعية علم النفس الأمريكية APA .

الجمهورية العربية السورية

وزارة التعليم العالي

جامعة تشرين



اسم الرسالة أو الأطروحة

(رسالة أعدت لنيل شهادة الماجستير أو الدكتوراه في اختصاص ....)

أعداد

الاسم الثلاثي للطالب

بإشراف

د. اسم المشرف (أو المشرفين)

الصفة العلمية

العام الدراسي

1- مخطط الرسالة أوالأطروحة:

وهكذا ويجب أن يتضمن كل فضل فقرتين على الأقل (ويتم ترتيب بقية الفصول في الفهرست كما تم الأمر بالنسبة للفصل الأول)

2- إرشادات عامة في الطباعة والإخراج:

- تتم الطباعة على ورق (297×210)A4 مع مراعاة الهوامش النظامية التالية:

من اليسار 2 سم، من اليمين 3 سم، من الأعلى 2 سم، من الأسفل 2 سم. ويكون شكل الصفحة كما هو مبين في هذا المنشور.

- يكتب رقم الصفحة في أسفل ومنتصف الصفحة كما يلي(صفحة x من y) بخط Simplified Arabic (12) مثل: (صفحة 1 من 155).
- يكتب عنوان الفقرة الرئيسة على يمين الصفحة دون مسافة بادئة بخط عربي عريض Simplified Arabic (18) وبخط أجنبي عريض (16) Times New Roman.
- يكتب عنوان الفقرة الفرعية على يمين الصفحة دون مسافة بادئة بخط عربي عريض Simplified Arabic (16) وبخط أجنبي عريض (14) Times New Roman.

- يكتب عنوان الفقرة الجزئية من الفقرة الفرعية على يمين الصفحة دون مسافة بادئة بخط عربي عريض Simplified Arabic (14) وبخط أجنبي عريض Times New Roman (12).
- يكتب نص الفقرة بخط عربي Simplified Arabic (13) وبخط أجنبي عريض Times New Roman (12) وذلك باستخدام (محاذاة- مضبوطة من الجانبين)، وباستخدام تباعد للخط مقداره 1.15 . ويبعد السطر الأول عن عنوان الفقرة (6 نقاط- وذلك باستخدام فقرة- تباعد من قائمة تنسيق في Microsoft Word كما يبعد عنوان الفقرة الأولى عن عنوان الفصل 18 نقطة، بينما تبعد نهاية الفقرة عن عنوان الفقرة التي تليها 6 نقاط إذا كانا في المستوى نفسه و12 نقطة إذا كانا في مستويين مختلفين.
- يبدأ السطر الأول لأية فقرة بمسافة بادئة تزيد بمقدار 1.27 سم عن بقية جسم النص لأي فقرة فرعية أو جزئية مضبوطاً من اليمين مع بداية رقم هذه الفقرة.
- وعند ورود كلمة أو عبارة أجنبية أو اختصاراً ضمن النص فترد مرة واحدة فقط مع ترجمتها إلى اللغة العربية وذلك حين ورودها لأول مرة ويكتفي بعد ذلك بالترجمة العربية أو الاختصار. ويتم الإشارة إلى ذلك في قائمة المصطلحات كما سيأتي توضيحه لاحقاً.

3- التعداد الرقمي والنقطي:

عند ورود تعداد رقمي و/أو نقطي في الفقرة تتم محاذة أرقام أو أحرف التعداد أو علامات الترقيم النقطي مع جسم الفقرة من جهة اليمين .

مثال (1):

1-1 : عنوان الفقرة الفرعية:

نص مكتوب يبدأ سطره الأول بمسافة بادئة تزيد بمقدار 0.6 سم عن بقية جسم النص مضبوطاً من اليمين مع بداية رقم الفقرة.

1- نص مكتوب نص مكتوب نص مكتوب نص مكتوب نص مكتوب نص مكتوب نص مكتوب نص مكتوب نص مكتوب نص مكتوب نص مكتوب.
2- نص مكتوب نص مكتوب نص مكتوب نص مكتوب نص مكتوب نص مكتوب نص مكتوب نص مكتوب نص مكتوب نص مكتوب نص مكتوب.
   أ. نص مكتوب نص مكتوب نص مكتوب.
   ب. نص مكتوب نص مكتوب نص مكتوب.
   ج. نص مكتوب نص مكتوب نص مكتوب.
   - نص مكتوب نص مكتوب نص مكتوب.
   - نص مكتوب نص مكتوب نص مكتوب.
   - نص مكتوب نص مكتوب نص مكتوب.

3- نص مكتوب نص مكتوب نص مكتوب نص مكتوب نص مكتوب نص مكتوب نص مكتوب.

4- طريقة كتابة المعادلات وترقيمها:

يراعى أن يتم توسيط المعادلات في الصفحة قدر الإمكان مع وضع رقم العلاقة الرياضية أو المعادلة على أقصى اليمين وبطريقة تدل على رقم المعادلة والفصل الموجود فيه.

مثال (2):

إن معادلة الإشارة $g_s(t)$ تعطى بالعلاقة:

$$g_s(t) = g(t) \sum_{n=-\infty}^{\infty} \delta(t - T_s) \qquad (2-3)$$

حيث: T هي (يتم هنا شرح دلالات الرموز المستخدمة)

ويمثل الترقيم (3-2) المعادلة رقم 2 في الفصل الثالث.

5- طريقة كتابة الأشكال وترقيمها:

يتم ترقيم وشرح الأشكال بخط غير بارز وبمقياس أصغر بقياس واحد من قياس النص الأصلي فإذا استخدمت للنص الأصلي قياس خط 12 تستخدم لشرح الأشكال وترقيمها قياس خط 11 مثلاً:

الشكل

الشكل (2-1): مخطط دارة بطاقة IBM النموذجية للإدخال والإخراج

وهذا يمثل الشكل الأول في الفصل الثاني

- **ترقيم الجداول**

إن ترقيم الجداول يختلف عن ترقيم الأشكال، حيث يتم وضع الرقم في أعلى الجدول وعلى يمينه كما يوضع المصدر في أسفل الجدول كما يلي:

الجدول (2-3): مكونات دارة بطاقة IBM النموذجية للإدخال والإخراج

المصدر: .....

- **طريقة الدلالة على استخدام المراجع ضمن النص:**

تتم الإشارة إلى المراجع المستخدمة في جسم النص بوضع رقم المرجع كما سيرد في قائمة المراجع وذلك ضمن قوسين مربعين إلى جانب العبارة أو الفقرة التي تم الرجوع فيها إلى المرجع المذكور.

مثال (3):

وقد أخذنا في اعتبارنا عند حل مسألة التصميم واستنتاج العلاقات الرياضية في هذا العمل تأثير ضجيج التداخل بين الرموز ISI كما ورد في كثير من الدراسات [7,8,9,10] عند دراسة الاستخدام الأمثل لعرض حزمة القناة.

ويدل ذلك على أن الدراسات المذكور رقم تسلسلها في قائمة المراجع قد تناولت دراسة تأثير ISI.

مثال (4):

يبين الشكل (1) المخطط العام لاستخدام التجميع بتقسيم التردد [11] من وهذا يشير إلى أن المخطط المذكور مأخوذ من المرجع [11].

6- تلخيص النتائج Results: (ماذا وجدت؟)

يعد هذا الجزء الأهم في الأطروحة. يجب أن تلخص المعطيات التي حصل عليها في هذا البحث من أجل ذلك يمكن إجراء ما يلي:

- تقديم النتائج بطريقة منطقية ومرتبة واستخدام التقسيمات نفسها التي وضعت في فقرة المواد والطرائق.
- التعبير عن النتائج على شكل جداول أو منحنيات بيانية . ولا تستخدم كلا الطريقتين للتعبير عن النتائج نفسها. وعندما يمكن كتابة النتائج ضمن متن النص فلا داعي لوضعها في جدول (في حال كانت النتائج فليلة).
- إعطاء أهم النتائج الهامة وليس المعطيات الخام التي تم الحصول عليها خلال البحث.
- يستعمل في السرد الزمن الماضي المبني للمجهول.

ومن الأمور التي يتوجب تجنبها في هذا القسم هي:

- إعطاء النتائج نفسها بأكثر من طريقة (جدول، شكل، ...).
- إهمال بعض المعطيات التي يعتقد أنها سلبية (بمعنى أنها لن تخدم فكرة المشروع).
- إعطاء المعطيات أو النتائج الخام الأولية.
- مناقشة النتائج أو إعطاء استنتاجات (هذا يجب أن يذكر في قسم المناقشة والخاتمة على التوالي).

7- عرض المناقشة Discussion: (ماذا تعني كل هذه النتائج)

من خلال هذا القسم يجب أن يتم:

- ربط النتائج بالنظريات أو الفرضيات الموضوعية: هل النتائج تبرهن على صحة النظرية أم لا ؟ ، كيف ولماذا ؟.
- مناقشة النتائج من خلال المشكلة أو الفرضية المطروحة في المقدمة.
- ربط النتائج بمسبباتها: أي لماذا حصلنا على هذه النتائج وكيف يمكن أن تكون ؟

- ربط النتائج بتلك التي حصل عليها باحثون آخرون عملوا على الموضوع نفسه أو موضوع مشابه له: هل النتائج تتفق معهم أم لا وما هو السبب؟ يجب شرح الأسباب وتجنب شرح النتائج بشكل واسع كما يجب تجنب الإسهاب والتعميم.

8- استخلاص الاستنتاجات Conclusions:

تجيب هذه الفقرة على التساؤلات الآتية:

- ما هي الاستنتاجات التي تم استخلاصها من النتائج ؟
- ما هي أهمية هذه النتائج بالنظر إلى المشكلة التي يتم العمل على حلها ؟
- ما هي أهم استخدامات هذه النتائج في تطبيقات عملية أو دراسات مستقبلية ؟

9- تقديم المقترحات أوالتوصيات Recommendations:

يتم ذكر أهم المقترحات التي يجب العمل بها والتوصيات التي يجب استكمال دراستها مستقبلاً للحصول على رؤية أشمل لحل المشكلة المدروسة.

10- طباعة المراجع References:

يجب ذكر جميع مصادر المعلومات والأفكار التي وردت في الرسالة أو الإطروحة، والإشارة إليها في متن النص وتثبيتها وترتيبها وترقيمها في قائمة المراجع وفق القواعد الآتية:

أولاً- في متن النص: يمكن الإشارة إلى المراجع ضمن النص بإحدى الطريقتين الآتيتين:

1- كتابة رقم المرجع حسب وروده في القائمة ضمن قوسين متوسطين كما يلي [1]، وذلك بعد أن يتم ترتيب المراجع وترقيمها وفق التسلسل في القائمة. ويمكن كتابة أرقام عدة مراجع ضمن القوسين ويتم الفصل فيما بينها بفواصل، مثل: [23,15,12]، وتوضع هذه الأرقام ضمن الأقواس بشكل متسلسل من الأصغر إلى الأكبر.

ثانياً: يوضع اسم الباحث (أي كنيته فقط بدون اسمه الأول وأيضاً كنية الثاني بدون اسمه الأول) مع عام النشر وذلك ضمن قوسين. وعندما يوجد عدة مراجع للمعلومة نفسها يتم كتابة هذه المراجع بحسب تسلسلها الزمني التصاعدي، أما في حال كان هناك مرجعين من العام نفسه فيتم ذكرهم بحسب تسلسل الأحرف الأبجدية. وفي حال كان المرجع قد أنجز من قبل أكثر من باحثين اثنين عندها يتم كتابة اسم الباحث الأول يليه كلمة "وآخرون" (إن كان المرجع عربياً) أو "et al." (إن كان المرجع أجنبياً على أن يتم كتابتها بشكل مائل، وأيضاً يجب كتابة اسم الباحث باللغة اللاتينية) .

ملاحظة: إن كلمة "et al." هي من كلمة et alii اللاتينية التي تعني "وآخرون" .

يمكن تلخيص ما سبق بالآتي:

1- مؤلف واحد: كتابة الكنية (باللغة الانكليزية في حال كان المرجع أجنبياً) متبوعاً بفاصلة ثم سنة النشر .

2- مؤلفان: كتابة كنية الأول مع كلمة "and" (في حال كان المرجع أجنبياً) ثم كنية المؤلف الثاني متبوعاً بفاصلة ثم عام النشر.
3- ثلاثة مؤلفين وما فوق: كتابة كنية الأول مع كلمة "وآخرون" أو ".et al" (في حال كان المرجع أجنبياً) متبوعاً بفاصلة ثم سنة النشر.
4- في حال ذكر أكثر من مرجع ضمن النص يتم تسلسل المراجع بحسب حداثتها من الأقدم إلى الأحدث مع وضع فاصلة منقوطة بين المرجع والآخر (إن كانت المراجع أجنبية).

أمثلة:

1- يتم تقييم طرائق الحفظ عن طريق الدراسات النموذجية في المنتج الغذائي (Leistner, 2000)
2- Phytoalexins هي مركبات ذات وزن جزئي منخفض تنتج من قبل النباتات الراقية رداً على الإصابة الميكروبية (Smid and Gorrhs, 1999).
3- استعملت التوابل قبل 1800 سنة لتحسين النكهة وتعديلها (Branen et al., 2001)
4- الوظيفة التقليدية لمضادات الأحياء الدقيقة الغذائية هي إطالة فترة الصلاحية (أو العمر التخزيني) وحفظ جودة المنتج الغذائي من خلال تثبيط نشاط الكائنات الحية الدقيقة المفسدة للأغذية.
(Branen et al., 2001; Davidson and Harrison, 2002; Davidson et al., 2005)

ويمكن استخدام الحالات الآتية في متن النص (بعض الأمثلة):

- يؤكد الباحث Jean (2001) أن نسبة المعادن في الخضار منخفضة.
- إن نتائج هذا البحث تتوافق مع ما وجده بعض الباحثين بأن نسبة المعادن في الخضار منخفضة (jean, 2001).

ملاحظة: لا يجب فصل التاريخ عن اسم العالم أو المؤلف.

ب_ في قائمة المراجع:

تتضمن قائمة المراجع الكتب والمقالات العلمية ومواقع الانترنيت وغيرها من المصادر الموثوقة، ويجب أن تشمل جميع المراجع المذكورة في متن النص وتراعى عند طباعتها القواعد التالية:

- كلمة "et al" المذكورة في متن النص يجب عدم كتابتها في صفحة المراجع بل يجب كتابة أسماء كل الباحثين المشاركين في هذا المرجع.
- في حال استخدام اسم الباحث في متن النص:

- يجب تدقيق كلا من أسماء الباحثين وسنة النشر كي تكون متطابقة مع تلك المذكورة في متن النص.
- يجب ترتيب المراجع في صفحة المراجع بحسب ترتيب الأحرف الأبجدية . وفي حال وجود مرجعين للباحث نفسه يرتبان زمنياً بشكل تصاعدي.

- في حال وجود مرجعين للباحث نفسه وفي العام نفسه، يوضع بجانب البحث الأول حرف a وبجانب الثاني حرف b .
- تقسم صفحة المراجع إلى مراجع عربية وأخرى أجنبية ولا تتضمن المراجع المحاضرات والنوط.

– في حال استخدام الأرقام في متن النص ترتب المراجع حسب ورودها في متن النص دون الفصل بين المراجع العربية والأجنبية.

يتم كتابة المراجع في صفحة المراجع وفق لما هو معمول به في مجلة جامعة تشرين للبحوث العلمية وتكتب المراجع كما يلي:

- إذا كان المرجع كتاباً أجنبياً:

الكنية بالأحرف الكبيرة، تتبعها فاصلة، الحرف الأول من الاسم تتبعه نقطة. الحرف الأول من الاسم المتوسط تتبعه نقطة. إذا تعدد المؤلفون، يفصل بين أسمائهم بفاصلة منقوطة (؛)، عنوان الكتاب أو البحث بالحرف المائل وتتبعه نقطة. الطبعة (ثانية، ثالثة ...) وتتبعها فاصلة، دار النشر وتتبعها فاصلة، بلد النشر وتتبعه فاصلة، سنة النشر وتتبعها فاصلة، عدد الصفحات وتتبعها نقطة.

مثال على ذلك:

TODD, D. K. Groundwater Hydrology. 2nd. ed.,John Willey & Sons, Inc New York & London, 1980,508.

- إذا كان المرجع كتاباً عربياً يتبع الأسلوب نفسه في كتابة المراجع الأجنبية، غير أن اسم الكاتب لا يختصر.
- إذا كان المرجع بحثاً منشوراً في مجلة باللغة الأجنبية.

يضاف بعد الكنية والاسم عنوان البحث بالحرف المائل متبوعاً بنقطة– اسم المجلة وبلد النشر وتتبعه فاصلة– المجلد والعدد (كتابة مختزلة) وبعدها فاصلة– وسنة النشر، أرقام الصفحات الخاصة بالبحث ضمن المجلة.

مثال على ذلك:

MAHOWALD, M. The Image of Jim the EHP Sequence. Annals of Mathematics U. S. A. Vol. 166, N^0. 1, 1982, 65-112 .

- إذا كان المرجع بحثاً منشوراً في مجلة باللغة العربية، يتبع الأسلوب نفسه في كتابة المراجع المنشورة في المجلات الأجنبية. غير أن اسم المجلة واسم الكاتب يكتبان من دون اختصار.

مثال على ذلك:

جواهري، محمد زهير. التدابير العلاجية في الآلام الرأسية، المجلة الطبية العربية . نقابة الأطباء في القطر العربي السوري، العدد السادس والثمانون، 1985، 54–59 .

- إذا كان المرجع من موقع إلكتروني، يضاف تاريخ المطالعة تتبعه نقطة، ثم يكتب العنوان الإلكتروني كاملاً بين الإشارتين .

AUSTEN, J. Pride and Prejudice, ed. Henry Churchyard, 1996, 10 Sept. 1998. http://www.pemberley.com/janeinfo/pridprej.html

أما في صفحة المراجع، فيبدأ باسم العائلة متبوعاً بفاصلة، تليها بقية الاسم متبوعة بنقطة، عنوان الكتاب متبوعاً بنقطة، مكان النشر متبوعاً بنقطتين، اسم الناشر متبوعاً بفاصلة، تاريخ النشر متبوعاً بنقطة. مثال:

Tannen, Deborah. You Just Don't Understand: Women and Men in Conversation. New York: Morrow, 1990 .

- وإذا كان المرجع مقالة في مجلة علمية يتبع الترتيب الآتي في الحاشية السفلية:

اسم الكاتب بالترتيب العادي متبوعاً بنقطة، عنوان المقالة متبوعاً بنقطة ضمن علامات تنصيص، عنوان المجلة بأحرف مائلة، رقم العدد متبوعاً بنقطة، رقم العدد، تاريخ النشر بين قوسين متبوعاً بنقطتين، ثم رقم الصفحة متبوعاً بنقطة. مثال:

Daniel C. Hallin. "Sound Bite News: Television Coverage of Elections, 1986-1988." Journal of Communication 42.2 (1992):5

أما في صفحة المراجع فيبدأ باسم العائلة متبوعا بفاصلة، بقية الاسم متبوعاً بنقطة، ثم التسلسل ذاته (كما في حالة الكتاب) مع تضمين أرقام صفحات المقال في المجلة. مثال:

Hallin, Daniel C. "Sound Bite News: Television Coverage of Elections, 1968-1988." Journal of Communication 42.2 (1992):5-24

ملاحظة هامة:

إذا كان المرجع بلغة غير الإنكليزية أو العربية مثل (الفرنسية أو الروسية أو الألمانية ...الخ) يكتب مضمون المرجع باللغة العربية أو اللغة الإنكليزية مع الإشارة إلى لغة المرجع .

[16] كيسيل، ف. ا. 1986– المصححات التشابهية والرقمية، دار راديو إسفياز، موسكو. (باللغة الروسية) .

Autran, J.C., 1995, Determination of common wheat in pasta products: an update of the achieved studies in the European (BCR) collaborative study 1990-1993 (Germany), Getreide Mehl und Brot, 49(5):272-277 .

11– طباعة الجداول والأشكال:

أ– الجداول:

- يهدف الجدول إلى تمركز كميات كبيرة من المعلومات واختصار في النص والكتابة.
- يجب أن لا يكون الجدول فائض عن النص.
- يستعمل الجدول أو تضمن المعلومة ضمن النص لكن لا تستخدم الاثنان معاً.
- يشار إلى الجدول ضمن النص بذكر أهم الأمور في الجدول دون الحديث عن كل ما يتضمنه الجدول.
- لا توضع معلومة في جدول إذا كان بالإمكان تضمينها بسهولة في النص.
- يجب أن يكون القارئ قادراً على فهم المعلومات في الجدول دون الرجوع إلى النص.

- يكتب عنوان الجدول فوق الجدول حكماً, ويلحق بواحدة القياس إن كانت موحدة وضرروربة.
- تحدد واحدات القياس بالطريقة الأكثر شيوعاً في رأس العمود أو في طرف الصف إن أمكن.
- عندما يتضمن الجدول نتائج على شكل متوسطات حسابية يجب عندها كتابة النتائج على الشكل: متوسط حسابي $\pm$ الانحراف المعياري $(mean \pm SD)$.
- يجب الإشارة إلى أرقام الجداول عند الاستفادة منها في متن النص.

ب- الأشكال:

يمكن اختيار الشكل المناسب وفق الآتي:

- المنحنيات البيانية Line graphs هي أكثر فعالية في إظهار المنحنى.
- الأشكال العمودية Bar charts هي أكثر فعالية لإظهار النسب النسبية (القيم النسبية).
- الخريطة المستديرة Pie charts هي أكثر فعالية لإظهار النسب من المجموع.
- الخرائط المشتركة Combined charts هي أكثر فعالية لإظهار الارتباطات.
- المخططات الإحصائية المتعلقة بالدراسة الإحصائية.

يجب مراعاة ما يلي عند استخدام الأشكال:

- يكتب عنوان الشكل دائماً في أسفله (وليس في أعلاه).
- تسمى المحاور الأفقية والعمودية مع ذكر وحدة القياس بين قوسين.
- يجب شرح الرموز المستخدمة ضمن الأشكال.
- يجب الإشارة إلى جميع أرقام الأشكال في متن النص.

12- تنسيق قائمة المصطلحات:

ترتب قائمة المصطلحات وفقاً للترتيب الأبجدي في اللغة الإنكليزية، وتطبع في عمودين متقابلين، بحيث يكون المصطلح الأجنبي على اليمين ويقابله المصطلح العربي كما هو مبين في القائمة التالية:

قائمة المصطلحات

باللغة العربية		باللغة الإنكليزية
	A	
تطبيقات		Application
	B	
لوحة		Board
مصرف		Bank
	C	
طاقة		Card
رمز		Code
⋮		⋮

وعند ورود اختصار لمصطلح ما ضمن قائمة المصطلحات فلابد من تفصيله مثلاً:

الشبكة المحلية LAN- local area network

ويفضل أن تحتوي الأطروحة دليلاً INDEX لورود الكلمات الهامة (المفتاحية) المستخدمة باللغة العربية ضمن الأطروحة حسب صفحات الورود . ودليلاً لورود الكلمات الهامة (المفتاحية) المستخدمة باللغة الإنكليزية ضمن الأطروحة حسب صفحات الورود ويرتب كل من الدليلين على عمودين كما في قائمة المصطلحات .

13- ترقيم الملاحق:

تسبق الملاحق ورقة خاصة يكتب عليها الملاحق، ويتم ترتيبها حسب الأحرف الأبجدية العربية ، مثلاً:

الملاحق:

الملحق (أ): عنوان الملحق

الملحق (ب): عنوان الملحق

الملحق (ج): عنوان الملحق

ويكتب رقم الملحق وعنوانه بخط مشابه لخط الفقرة الرئيسة كما هو مبين أعلاه, كما يمكن أن تذكر تفصيلات الملاحق، أن كانت رسوماً أو جداول أو نصاً مكتوباً أو صورة أو غير ذلك . ويمكن أن يستهلك الملحق الواحد صفحة واحدة أو عدة صفحات.

14- ملاحظات مهمة:

1- تستخدم الأرقام العربية (1,2,3...) حصراً في كافة صفحات المشروع.

2- في حال كان المشروع يتضمن دراسة لخط إنتاجي، فيجب أن بتضمن المشروع مخططاً تكنولوجياً موضحاً عليه كل المعلومات وعلى ورق تيراج.

3-يتم تصحيح نسخة المشروع المسلمة للمكتبة بعد جلسة التحكيم وفقا لملاحظات اللجنة وتعتمد النسخة المصححة من قبل رئيس القسم بعد تأشير الدكتور المشرف عليها، وذلك من أجل الحصول على براءة الذمة من المكتبة.

15- تذكرة بعلامات الترقيم وكيفية استخدامها:

وهي علامات ضرورية لما لها من تأثير على فهم المعنى والسياق والترابط بين الجمل والعبارات. والنص الخالي من علامات الترقيم نص أبكم أصم، قد يحمل القارئ على إعادته مرتين أو أكثر ليفهم محتواه، أما النص المتضمن علامات الترقيم بشكل صحيح، فهو نص ناطق مبين، يُظهر أفكاره للقارئ فكرة فكرة، ويشرح مقاصده ومعانيه، ويضع أصبعه وعلى كل مافيه من عاطفة وتنوع في التعبير.

لذلك سنُذكر فيما يلي بكيفية استخدام علامات الترقيم في الكتابة باللغة العربية. وعلامات الترقيم هى:

1- النقطة (.): وتوضع في نهاية الجملة التامة المعنى وكذلك عند انتهاء الكلام.

2- الفاصلة (،): وتوضع في الأحوال التالية:

- بعد لفظ المنادى، مثل: يا حسن، أحضر الكتاب.
- بعد الجملتين المرتبطتين بالمعنى والإعراب، مثل:خير الكلام ماقل ودل، وشره ماطال وذل. الربيع جميلٌ، وأزهاره فواحة. أقبل المزارع إلى المزرعة، وأقبل معه أولاده.
- بين الشرط والجزاء وبين القسم والجواب، مثل:

إذا كنت في مصر ولم تكن ساكناً على نيلها الجاري، فما أنت في مصر.

- بين المفردات المعطوفة في الجمل الطويلة، مثل:

ما خاب تاجر صادق، ولا تلميذ عمل بنصائح والديه ومعلميه.

3- الفاصلة المنقوطة (؛): وتوضع في الحوال التالية:

- بعد جملة ما بعدها سبب أو تفسير لها، مثل:

محمد أفضل الطلاب؛ لأنه أكثرهم جداً، وأفضلهم خلقاً. أذاكر دروسي؛ لأنجح في حياتي.

- بين جملتين مرتبطتين في المعنى دون الإعراب، مثل:

إذا رأيتم الخير فخذو به؛ وإذا رأيتم الشر فدعوه.

4- النقطتان (:): وتوضعان في الأحوال التالية:

- بين القول والمقول، مثل:

قال رسول الله (ص): علموا أولادكم الرماية والسباحة وركوب الخيل.

- بين الشيء وأقسامه، مثل:

يحتوي الوطن العربي على الأقطار التالية: سوريا- مصر- العراق ...الخ .

- قبل الأمثلة: نمارس بعض الهوايات، مثل: الرياضة والموسيقى والمطالعة.

5- علامة الاستفهام (؟): وتوضع بعد جملة الاستفهام، مثل: كم عدد الطلاب في صفك ؟

6 - علامة التعجب (!): وتوضع بعد كل جملة يعبر فيها عن التعجب والدهشة، مثل: ماأجمله!.

مراجع الجزء الأول

المراجع باللغة العربية:

1- البهي، فؤاد السيد. (1979)- علم النفس الإحصائي. ط3، دار الفكر العربي، عين شمس، مصر.
2- جون د. + يكنسون ب. (1987)- العلم والمستقلون بالبحث العلمي. عالم المعرفة- الكويت.
3- التير، مصطفى عمر. (1989)- مساهمات في أسس البحث الاجتماعي. معهد الإنماء العربي والدراسات الاجتماعية، ليبيا.
4- حسن، عبد الباسط محمد. (1990)- أصول البحث العلمي. مكتبة وهبي، القاهرة، مصر.
5- زايد، مصطفى. (1984)- الإحصاء ووصف البيانات. دار العلوم، الرياض، السعودية.
6- سيد أحمد، غريب محمد. (1985)- الإحصاء والقياس في البحث الاجتماعي. دار المعرفة الجامعية، الاسكندرية، مصر .
7- العلي، ابراهيم محمد. (2002)- مبادئ علم الإحصاء. جامعة تشرين، اللاذقية، سوريا .
8- العلي، ابراهيم محمد+ كابوس، أمل. (1986)- الإحصاء الرياضي. جامعة حلب، حلب، سوريا.
9- عبيدات ذوقان وآخرون. بلا تاريخ- البحث العلمي. دار مجدلاوي، عمان، الأردن.
10- عمر، محمد زياد. (1981)- البحث العلمي. دار الشروق، الرياض، السعودية.
11- غرايبة، فوزي وآخرون. (1987)- أساليب البحث العلمي. الجامعة الأردنية، عمان، الأردن.

المراجع باللغة الأجنبية:

1- Adams, J. Khan, H. Raeside, R. and White, D. (2007). Reseach Methods for Graduate and Social Science Students. Response Book, Sage puplications .
2- Beins, B. C. and Mc carthy, M. A. (2012), Research Methods and Statistics. Pearson Education.
3- Cooper, D. R. and Schindler, P. S. (2014), Business Research Methods. 6th ed. Mc Grow Hill.
4- Fisher, C. (2007), Researching and Writing a Dissertation for Business Students. 2th ed. Harlow: Financial Times Prentice Hall.
5- Foddy, W. (1994). Constracting Questions for Interviews and Questionnaires. Combridge University Press.
6- Sannders, L. Lewis, P. and Thornhill, A. (2009). Research Methods for Buscness Students. 5th ed. Prentice Hall.

الجـزء الثاني

أساليب التحليل الإحصائي في البحث العلمي

يتناول هذا الجزء بعض الأساليب الإحصائية التي يمكن أن تستخدم في معالجة وتحليل المعلومات الإحصائية المتوفرة عن الظواهر المدروسة في مشاريع البحث العلمي. ونقصد بالمعلومات القيم العددية (البيانات) أو الحالات الوصفية (الصفات) , التي تعبر عن المتحولات المعرفة على الظاهرة المدروسة . وبذلك ويمكننا تصنيف المعلومات إلى نوعين أساسيين هما:

أ- معلومات كمية: وهي بيانات عددية عن متحولات قابلة للقياس بواحدات قياس محددة، وهذه البيانات يمكن أن تكون:

- منقطعة: كعدد أفراد الأسرة- وعدد الطلاب- وعدد السيارات ...الخ.
- مستمرة: كعمر الإنسان- درجة الحرارة- مقدار الدخل ....الخ.

ب- معلومات نوعية: وهي حالات وصفية لمتحولات غير قابلة للقياس، وهذه المعلومات يمكن أن تكون:

- أسمية: كحالات الجنس- حالات العمل- الحالة الاجتماعية ...الخ.
- مرتبة: كحالات التعليم- حالات الوظيفة- حالات الرضا ..الخ.

ويتم جمع المعلومات عن الظاهرة المدروسة أو عن المتحولات المطلوبة من عناصر المجتمع الاحصائي بواسطة أحد الأسلوبين:

- الحصر الشامل: وهو يشمل جميع عناصر المجتمع الاحصائي المؤلف من N عنصراً .
- المسح بالعينة: وهو يشمل جزء من المجتمع ويكون على شكل عينة حجمها n عنصراً, تسحب عشوائياً من عناصر ذلك المجتمع بدون إعادة أو مع الاعادة .

وتستخدم بيانات هذه العينة لتقدير معالم المجتمع المجهولة مثل: المتوسط $\boldsymbol{\mu}$ أو التباين $\boldsymbol{\sigma^2}$ أو نسبة خاصة فيه $\boldsymbol{R}$، وذلك من خلال استخدام المؤشرات المقابلة لها والمحسوبة من العينة، والتي سنسميها (مؤشرات العينة)، وهي: متوسط العينة $\overline{\boldsymbol{x}}$ وتباين العينة المصحح $\boldsymbol{S^2}$ ونسبة تلك الخاصة في العينة $\boldsymbol{r}$. ويجب أن تكون مؤشرات العينة تقديرات غير متحيزة وفعالة ومتماسكة لمعالم المجتمع المقابلة لها , لذلك يتم تعديلها أو تصحيحها حتى تحقق تلك الشروط .

ويتضمن هذا الجزء الفصول التالية:

الفصل الأول: أساليب معالجة وعرض المعلومات الإحصائية.

الفصل الثاني: حساب مقاييس النزعة المركزية والتشتت.

الفصل الثالث: المتحولات العشوائية والتوزيعات الاحتمالية.

الفصل الرابع: العينات ومسائل التقدير.

الفصل الخامس: تصميم وتحليل الاستبيان.

الفصل السادس: اختبارات الفرضيات البسيطة.

الفصل السابع: تحليل التباين البسيط.

الفصل الثامن: الارتباط والانحدار والسلاسل الزمنية.

الفصل التاسع: التحليل اللوجستي.

الفصل العاشر: الاختبارات اللامعلمية.

وهناك عدة أساليب أخرى لم نستطع التعرض لها في هذه المحاضرات, لكونها تحتاج إلى دراسات موسعة وبراهين معقدة, مثل: التحليل العاملي, التحليل التمييزي، تحليل الارتباط القانوني, التحليل المتقدم والمتعدد للانحدار,التحليل المتقدم للسلاسل الزمنية, تحليل الشبكات العصبونية...الخ.

علماً بأن بعض هذه الأساليب موجودة في كتابنا: أسس التحليل الإحصائي متعدد المتغيرات, المنشور إلكترونياً على عدة مواقع علمية.

الفصل الأول
أساليب معالجة وعرض المعلومات الإحصائية

يتناول هذا الفصل بعض أساليب معالجة المعلومات الإحصائية وعرضها، وهي: الترتيب، التبويب، حساب التكرارات المطلقة والنسبية، حساب التكرارات التجميعية (المتصاعدة أو المتنازلة)، التمثيل البياني، إنشاء المنحنيات التكرارية.

1-1: ترتيب المعلومات الإحصائية:

ويُقصد به تنظيم المعلومات المفردة في جدول مناسب ترتيباً تصاعدياً أو تنازلياً، وحساب التكرارات المقابلة للقيم الأساسية المرتبة.

مثال (1-1): لنأخذ المعلومات الافتراضية الآتية عن أعمار (44) طالباً في الجامعة, ونقرؤها ونسجلها كما وردت في الاستمارة الإحصائية لكلٍ منهم (وهي مقاسة بالسنوات الصحيحة):

22	20	21	23	24	22	25	24	21	22	20
22	23	25	21	20	21	22	25	23	24	23
22	21	23	20	24	22	25	21	24	23	21
22	23	25	24	22	23	21	22	23	21	24

من خلال قراءة هذه المعلومات نلاحظ أنها مؤلفة من (44) قيمة عددية، مبعثرة بشكلٍ عشوائي، ولا تفيدنا كثيراً في استخلاص أية نتيجة عن عمر الطالب في الجامعة، لذلك نحاول أن نرتبها تصاعدياً، ونحسب التكرارات المطلقة المقابلة لكل قيمة منها، ونتبع الخطوات الآتية:

1-1-1: خطوات الترتيب:

1 . نبحث عن أصغر قيمة بينها فنجد أنها القيمة (20)، كما نحدد أكبر قيمة فنجدها (25).

2 . نعدّ جدولاً مناسباً، ونضع في سطره (أو عموده) الأول القيم العددية الأساسية مرتبةً ترتيباً تصاعدياً ابتداءً من القيمة الصغرى (20)، وحتى القيمة الكبرى (25)، فنحصل على السطر الأول في الجدول (1-1) المبيّن أدناه.

3 . نقوم بحساب تكرارات كل قيمة من القيم المرتبة، وذلك بإجراء مسح أو قراءة لكل تلك الأعداد، ثم نضع إشارة خط عمودي أمام القيمة كلما وردت خلال القراءة، ولتسهيل العمل والحساب، ندمج كل خمس إشارات بإشارة خامسة أفقية أو مائلة، كما في الجدول (1-1).

4 . نحسب التكرارات المقابلة لتلك القيم، ونسجلها في سطر خاص بعنوان (التكرارات المطلقة)، ثم نتأكد من أن مجموع تلك التكرارات يساوي عدد القياسات المفردة وهو (44) طالباً، وبذلك نحصل على عناصر السطر الثالث من الجدول (1.1) الآتي:

جدول (1-1): عمر الطالب في الجامعة لعام 2009 (العدد الكلي 44 طالباً).

العمر (بالسنوات)	20	21	22	23	24	25	المجموع
إشارات القراءة	IIII	𝍸 IIII	𝍸 𝍸	𝍸 IIII	𝍸 II	𝍸	
التكرارات المطلقة	4	9	10	9	7	5	44

المصدر: بيانات المثال(1-1).

وبذلك نكون قد قمنا بترتيب أعمار الطلاب ترتيباً تصاعدياً، قد حصلنا على جدول مختصر يتضمن جميع المعلومات المفردة السابقة، ويفيدنا هذا الجدول كثيراً في استخلاص العديد من النتائج الهامة مثل:

- إن أعمار الطلاب محصورة بين 20 و25 عاماً (حسب المثال المفروض).
- إن أكثر الأعمار تكراراً هو العمر 22 , لأنه يقابل أكبر تكرار (10 مرات)، وتسمى القيمة الأكثر تكراراً بالمنوال، وهو أحد مقاييس النزعة المركزية.
- إن عدد القيم العددية المفردة هو (44) قيمة، بينما عدد القيم الأساسية المرتبة هو (6) قيم فقط، وهي: 25، 24، 23، 22، 21، 20.
- يمكننا حساب التكرارات النسبية لكل من الأعمار السابقة, وذلك بتقسيم كل من التكرارات المطلقة على مجموعها (44) طالباً, فنحصل على ما يسمى بالتوزيع التكراري لأعمار الطلاب المدروسين, والذي يتألف من أعداد كسرية غير سالبة، وأن مجموعها يساوي الواحد.

جدول (1-2): التوزيع التكراري لأعمار الطلاب في المثال(1-1).

العمر (بالسنوات)	20	21	22	23	24	25	المجموع
التكرارات المطلقة	4	9	10	9	7	5	44
التكرارات النسبية أو التوزيع التكراري	0.09	0.205	0.23	0.205	0.16	0.11	1
التكرارات التجميعية المتصاعدة (التوزيع التجميعي)	0.09	0.295	0.525	0.730	0.890	1	

- ومنه يمكننا حساب الاحتمالات المنفردة من قيم التوزيع التكراري، فإذا رمزنا لعمر الطالب بـ A، ولقيمة الاحتمال بالرمز P، فنجد من الجدول السابق أن احتمال أن يكون عمر الطالب A مساوياً لـ (20) عاماً يساوي 0.09، ونكتب ذلك كما يلي:

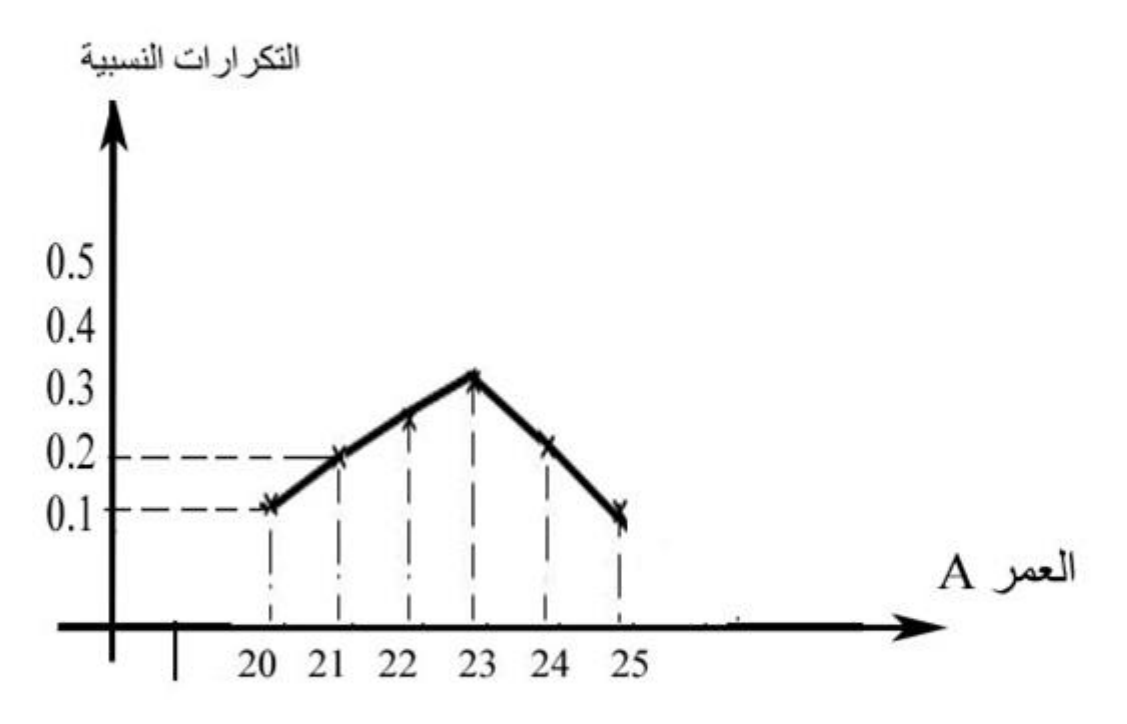
$P(A=20) = 0.09$

وهكذا نحصل على بقية الاحتمالات فنجد أنها تساوي:

$P(A=21) = 0.205$
$P(A=22) = 0.23$
$P(A=23) = 0.205$
$P(A=24) = 0.16$
$P(A=25) = 0.11$

وهنا نلاحظ أن مجموع هذه الاحتمالات يساوي الواحد تماماً.

– ويمكننا تمثيل هذا التوزيع التكراري (التكرارات النسبية) بيانياً على المحورين الإحداثيين، فنضع قيم الأعمار على المحور الأفقي، ونضع قيم التوزيع التكراري على المحور العمودي، ثم نرسم النقاط الهندسية لكل قيمة عمرية مع التكرار النسبي المقابل لها، فنحصل على الشكل الآتي:



الشكل (1–1): التوزيع التكراري لأعمار الطلاب أو المضلّع التكراري

– كما يمكننا حساب التكرارات التجميعية المتصاعدة المقابلة للقيم المرتبة, وذلك بإضافة كل تكرار نسبي إلى سوابقه, فنحصل على ما يسمى بالتوزيع التجميعي المتصاعد، وهو يلعب دوراً كبيراً في حساب الاحتمالات التراكمية التي يكون لها شكل أصغر أو تساوي (≤) .

فمثلاً، نجد أن احتمال أن يكون عمر الطالب أصغر أو يساوي (23) عاماً يساوي:

$P(A\leq 23) = P(A=20)+ P(A=21)+ P(A=22)+ P(A=23)$
$P(A\leq 23) = 0.09 + 0.205+ 0.23+ 0.205$
$P(A\leq 23) = 0.730$

وهنا نشير إلى أنه تم حساب هذا الاحتمال سابقاً ووضعه في الجدول السابق في سطر التكرارات التجميعية المتصاعدة مقابل العمر (23).

وهكذا يمكننا وبسرعة حساب قيم جميع الاحتمالات التجميعية المتراكمة المتصاعدة من خلال سطر التكرارات التجميعية المتصاعدة، حيث نجد أن:

احتمال أن يكون عمر الطالب أصغر أو يساوي 20 : $P(A \leq 20) = 0.09$

احتمال أن يكون عمر الطالب أصغر أو يساوي 21 : $P(A \leq 21) = 0.295$

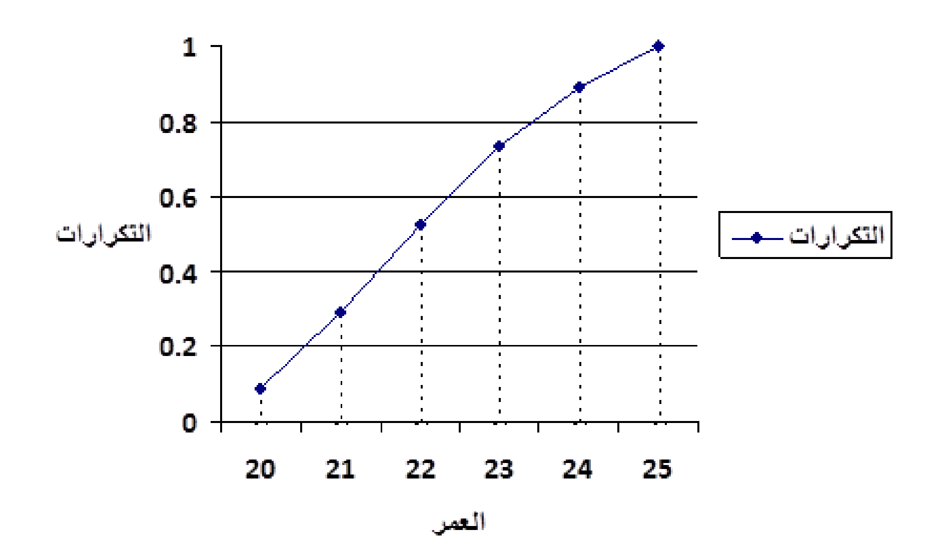
احتمال أن يكون عمر الطالب أصغر أو يساوي 22 : $P(A \leq 22) = 0.525$

احتمال أن يكون عمر الطالب أصغر أو يساوي 23 : $P(A \leq 23) = 0.730$

احتمال أن يكون عمر الطالب أصغر أو يساوي 24 : $P(A \leq 24) = 0.890$

احتمال أن يكون عمر الطالب أصغر أو يساوي 25 : $P(A \leq 25) = 1$

– ويمكننا تمثيل التكرارات التجميعية المتصاعدة بيانياً, فنجد أنها تأخذ الشكل الآتي:



الشكل (1–2): التكرارات التجميعية المتصاعدة

– ويمكننا حساب النسب المئوية لأعمار الطلاب، وذلك بتقسيم التكرارات المطلقة المقابلة لكل فئة عمرية على مجموعها (44)، ثم ضرب الناتج بـ 100 ووضع إشارة % بعد حاصل القسمة للدلالة على أنها نسبة مئوية، فنحصل على الجدول الآتي:

جدول (1–3): النسب المئوية لأعمار الطلاب في المثال(1–1).

المجموع	25	24	23	22	21	20	العمر (بالسنوات)
44	5	7	9	10	9	4	التكرارات المطلقة
100	11	16	20.5	23	20.5	9	النسبة المئوية %

وهنا نلاحظ أن النسب المئوية ما هي إلا قيم التوزيع التكراري مضروبةً بـ 100, ويُستفاد من النسب المئوية في إظهار حجم أو ثقل كل من الأعمار السابقة بين مجموعة الأعمار الأخرى، حيث نجد أن النسبة المئوية للطلاب الذين تبلغ أعمارهم 22 عاماً تساوي 23% .... إلخ.

– يمكننا تمثيل أو عرض النسب المئوية بيانياً ضمن دائرة واحدة، بحيث نخصص لكل عمر قطاعاً معيناً منها، يقابل زاوية تتناسب مع قيمة النسبة المئوية المقابلة له.

لذلك نقوم بحساب الزاوية التي تتناسب مع كل نسبة مئوية، فنفترض أن محيط الدائرة , الذي يساوي 360 درجة يقابل 100 جزءاً، ونحسب مقدار الزاوية المقابلة للنسبة الأولى كما يلي:

إن كامل محيط الدائرة 360 درجة، يقابل 100 جزء.

وإن زاوية القطاع الأول A1 للنسبة الأولى ، تقابل 9 أجزاء من المائة.

وبذلك نجد أن زاوية القطاع الأول المقابل للعمر (20) عاماً تساوي:

$$\boldsymbol{A_1 = \frac{360 \cdot 9}{100} = 32.40}$$ درجة

وهكذا نحصل على زاوية القطاع الثاني المقابل للعمر (21) عاماً تساوي:

$$\boldsymbol{A_2 = \frac{360 \cdot 20.5}{100} = 73.80}$$ درجة

وزاوية القطاع الثالث المقابل للعمر (22) عاماً تساوي:

$$\boldsymbol{A_3 = \frac{360 \cdot 23}{100} = 82.80}$$ درجة

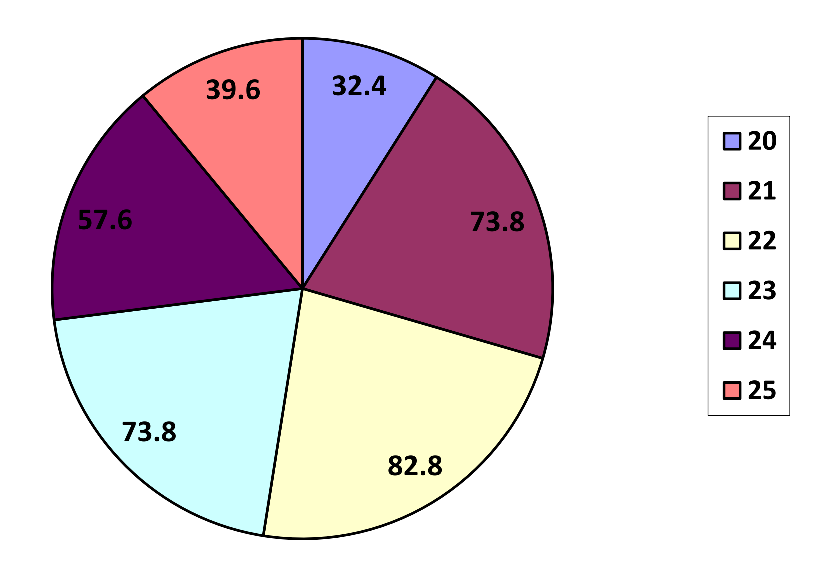
وهكذا نحسب بقية الزوايا القطاعية للأعمال 23، 24 و25، فنجد أن:

$$\boldsymbol{A_4 = \frac{360 \cdot 20.5}{100} = 73.80}$$ درجة

$$\boldsymbol{A_5 = \frac{360 \cdot 16}{100} = 57.60}$$ درجة

$$\boldsymbol{A_6 = \frac{360 \cdot 11}{100} = 39.60}$$ درجة

ثم نقوم يدوياً أو بواسطة الحاسوب برسم الدائرة المناسبة ونرسم بداخلها القطاعات الزاوية المذكورة فنحصل على الشكل الآتي:



الشكل (1−3): تمثيل النسب المئوية لأعمار الطلبة

مثال (1−2):

لنأخذ مثالاً عن الحالة التعليمية لسكان قرية معينة، ولنفترض أن الدراسة الميدانية للسكان, الذين تزيد أعمارهم عن 10 سنوات أظهرت أن عددهم الكلي يبلغ (3000) نسمة. ومن عملية تصنيف هؤلاء السكان حسب الحالة التعليمية وحساب التكرارات المقابلة لكل حالة، ثم ترتيبها تصاعدياً في جدول مناسب، حصلنا على الجدول الآتي:

جدول (1−4): النسب المئوية لأعمار الطلاب.

المجموع	عليا	جامعي	متوسط	ثانوي	إعدادي	ابتدائي	متعلم	أمي	الحالة التعليمية
3000	50	150	300	800	700	500	300	200	**التكرارات المطلقة**
1	0.016	0.050	0.100	0.267	0.233	0.167	0.100	0.067	**التكرارات النسبية التوزيع التكراري**
×	1	0.984	0.934	0.834	0.567	0.334	0.167	0.067	**التوزيع التجميعي المتصاعد**
100	1.6	5.0	10.0	26.7	23.2	16.7	10.0	6.7	**النسبة المئوية المقابلة %**

المصدر: فرضي.

ومن خلال ترتيب هذه المعلومات النوعية يمكننا أن نستنتج الكثير من النتائج، مثل:

. إن أكبر تكرار يقابل حالة الشهادة الثانوية، حيث ظهر لدينا 800 تكراراً لها من أصل 3000 نسمة (إذن الثانوية هي المنوال).

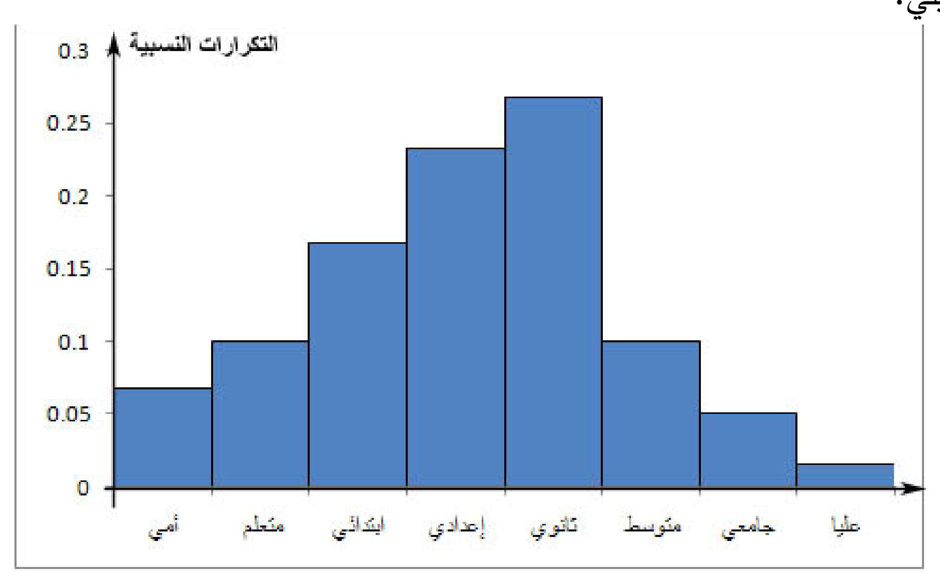
. يمكننا حساب الاحتمالات المنفردة والتجميعية كما فعلنا في المثال السابق، حيث نجد أن:

احتمال أن يكون الشخص المختار عشوائياً يحمل ابتدائية = 0.167

احتمال أن يكون الشخص المختار عشوائياً يحمل شهادة متوسطة = 0.10.

وكذلك نجد أن احتمال أن تكون شهادة الشخص المختار عشوائياً أقل من ثانوية = 0.567.

وكذلك نجد أن نسبة الأمية تساوي 6.7%، وأن نسبة الجامعة 5% فقط. وأخيراً يمكننا تمثيل التوزيع التكراري لهذه الحالات بيانياً على المحورين الإحداثيين على شكل أعمدة فوق كل حالة على المحور الأفقي كما يلي:



الشكل (1-4): التمثيل البياني التوزيع التكراري.

كما يمكننا تمثيل النسب المئوية لهذا التوزيع التكراري على شكل دائرة وضمنها قطاعات زاوية مقابلة لكل حالة حسب نسبتها المئوية، فنحصل على الشكل الآتي:



الشكل (5-5): تمثيل النسب المئوية للحالة التعليمية.

ملاحظة: نترك تفاصيل الحسابات للطالب ليتأكد منها، وإعادة رسم التمثيل البياني بدائرة جديدة.

1-2: تبويب المعلومات الإحصائية:

عندما تكون المعلومات الإحصائية الكمية ناتجة عن متحولات مستمرة أو عندما يكون عدد المعلومات المنقطعة كبير جداً (كعدد السكان)، يُفضل أن نلجأ إلى عملية التبويب أو التصنيف بدلاً من الترتيب:

التبويب: هو تجميع المعلومات ضمن مجالات جزئية أو فئات نوعية وحساب التكرارات المقابلة لكل منها. وهنا نشير إلى ضرورة أن تكون تلك المجالات أو الفئات متلاصقة وغير متقاطعة فيما بينها، وذلك حتى لا يقع التباس في عملية حساب التكرارات.

مثال(1-3): لنأخذ عدد سكان سورية حسب تعداد عام 2004، فنجد أنه بلغ (17921) ألف نسمة. فإذا أردنا ترتيبهم حسب العمر, لتطلب منا الأمر تنظيم جدول فيه أكثر من 100 سطر، لذلك نلجأ إلى تبويبهم حسب العمر والنوع ضمن فئات عمرية خمسية كما يلي:

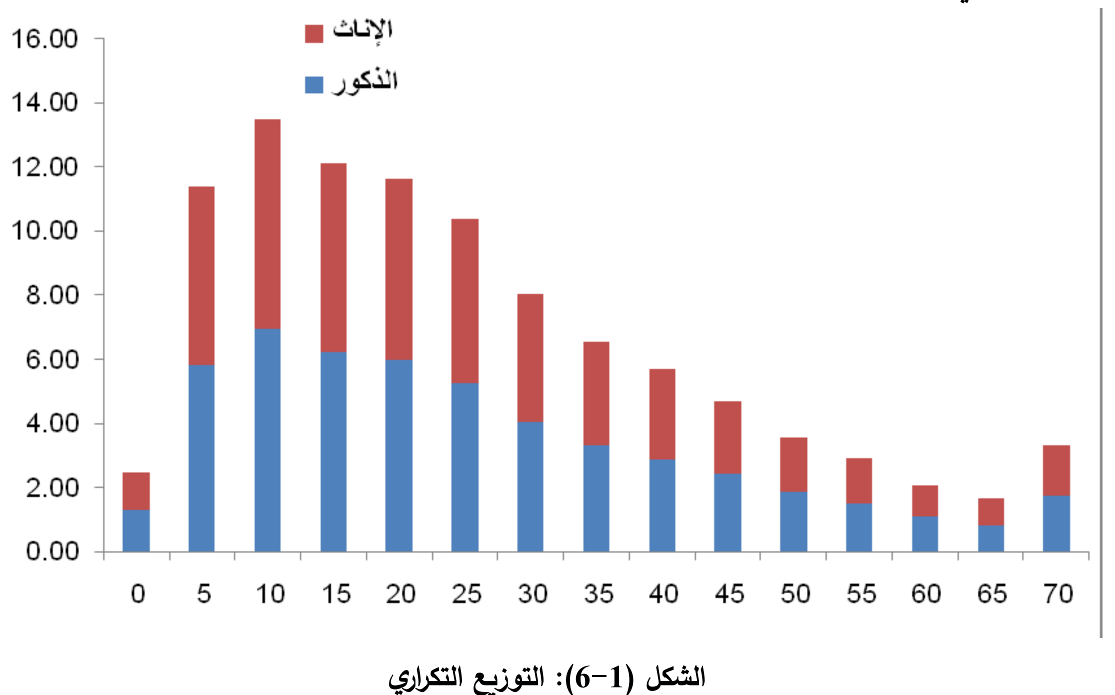
جدول (1-5): تبويب سكان سورية عام 2004 حسب العمر والنوع (ألف نسمة)

فئات العمر	عدد الذكور	عدد الإناث	المجموع	التوزيع النسبي للذكور%	التوزيع النسبي للإناث%	التوزيع النسبي للمجموع%
أقل من سنة	229	210	439	...	...	2.5
1-4	1044	999	2043	...	...	11.4
5-9	1246	1174	2420	...	...	13.5
10-14	1118	1051	2169	...	...	12.1
15-19	1072	1015	2097	...	...	11.2
20-24	944	920	1864	...	...	10.4
25-29	724	718	1442	...	...	8.1
30-34	595	578	1173	...	...	6.6
35-39	513	508	1021	...	...	5.7
40-44	431	412	843	...	...	4.7
45-49	330	307	637	...	...	3.5
50-54	266	254	520	...	...	2.9
55-59	192	175	367	...	...	2.0
60-64	146	149	295	...	...	1.6
65 فأكثر	311	280	591	...	...	3.3
المجموع (ألف)	9161	8760	17921	**51.1**	**48.9**	100

المصدر: المجموعة الإحصائية لعام 2006، ص 64+65.

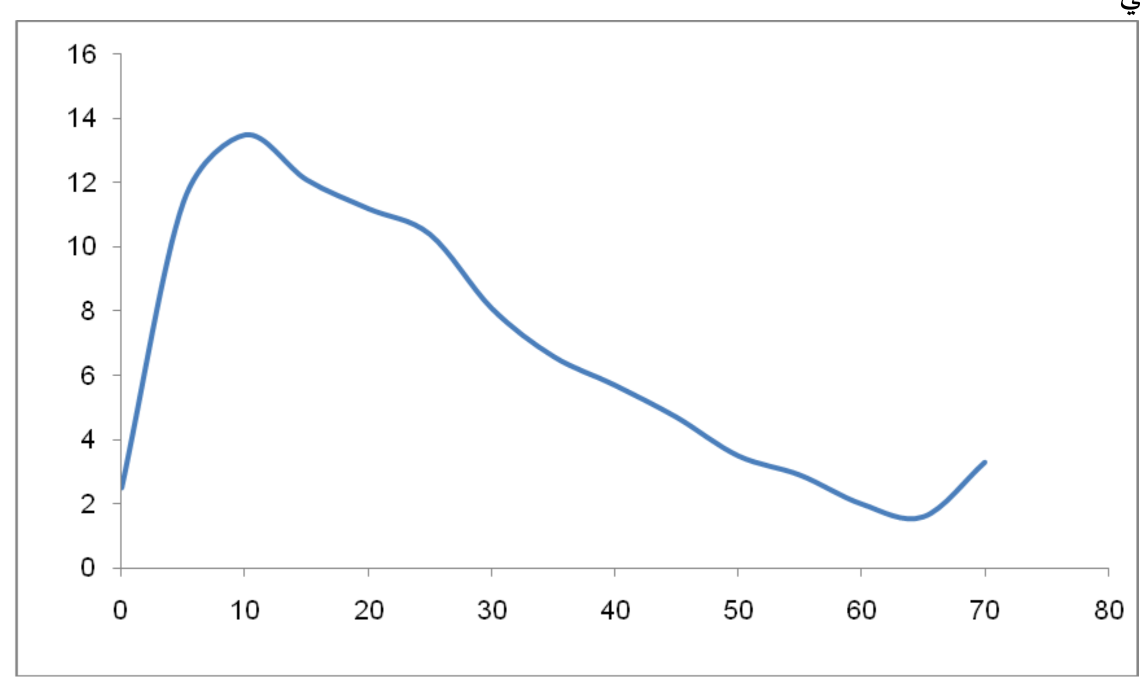
ونترك للطالب حساب نسب الذكور والإناث في كل فئة عمرية وذلك بتقسيم عدديهما على عدد السكان 17921.

ويمكننا تمثيل التوزيع التكراري لمجموع عدد السكان (التوزيع النسبي لمجموع الذكور والإناث)، على المحورين الإحداثيين، فنضع العمر على المحور الأفقي والتوزيع النسبي على المحور العمودي فنحصل على الشكل التالي, الذي يظهر التمثيل على شكل أعمدة ونسبة فوق كل فئة عمرية، وإن كل عمود يتألف من جزأين السفلي للذكور والعلوي للإناث.



الشكل (1-6): التوزيع التكراري

وإذا قمنا بوصل منتصفات رؤوس هذه الأعمدة بخط انسيابي لحصلنا على منحني مستمر يأخذ الشكل التالي:



الشكل (1-7): المنحني التكراري لأعمار السكان في سورية

1-2-1: خطوات التبويب:

1 . تحديد نوع التبويب المناسب للظاهرة المدروسة والذي يخدم الهدف المنشود من البحث.

2 . تصميم الجدول المناسب لعملية التبويب وتحديد عنوانه وأسماء أعمدته وأسطره.

3 . تحديد عدد الفئات أو المجالات المستخدمة في التبويب، ويمكن الاستفادة من علاقة struges التالية:

$$m = [1 + 3.322 \lg_{10} n]$$

4 . تحديد أو حساب أطوال الفئات والمجالات الجزئية d من العلاقة:

$$d = \frac{X_{\max} - X_{\min}}{m} = \frac{X_{\max} - X_{\min}}{[1 + 3.322 \lg_{10} n]}$$

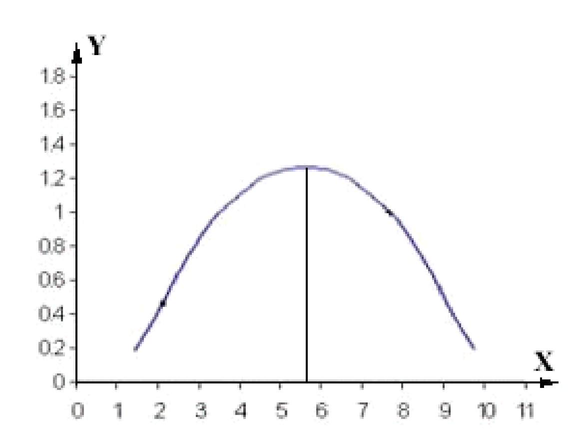
5 . حساب التكرارات المقابلة لكل فئة أو مجال جزئي وتثبيتها في الجدول مقابل تلك الفئة.

6 . حساب مجموع التكرارات والتأكد من أنه يساوي المجموع الكلي n.

1-3: أشكال منحنيات التوزيعات التكرارية:

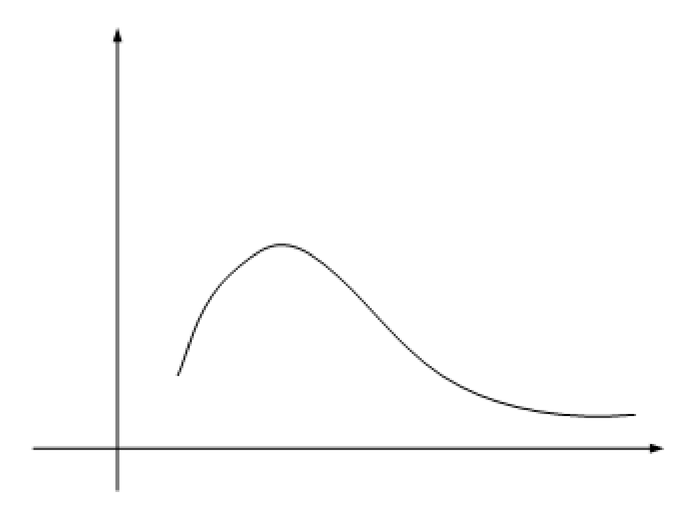
إن منحنيات التوزيعات التكرارية تأخذ أشكالاً مختلفة، وذلك تبعاً لطبيعة الظاهرة المدروسة، ولكنه يمكننا تصنيف هذه المنحنيات إلى عدة أشكال مميزة هي:

1 . المنحنيات المتناظرة: وهي التي تكون محدبة أو مقعرة من الوسط ومتناظرة على الطرفين بالنسبة للمتوسط مثل :



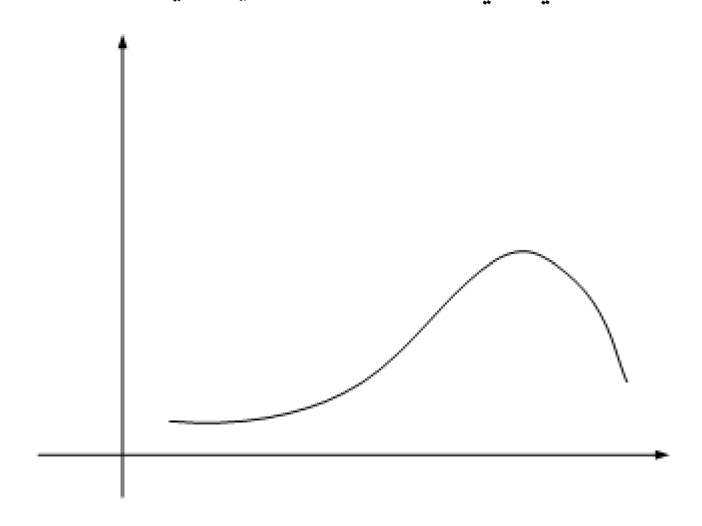
الشكل (1-8): المنحنيات المتناظرة

2 . المنحنيات المائلة إلى اليمين: وهي التي يكون طرفها اليميني طويلاً بالنسبة لمركزها، مثل:



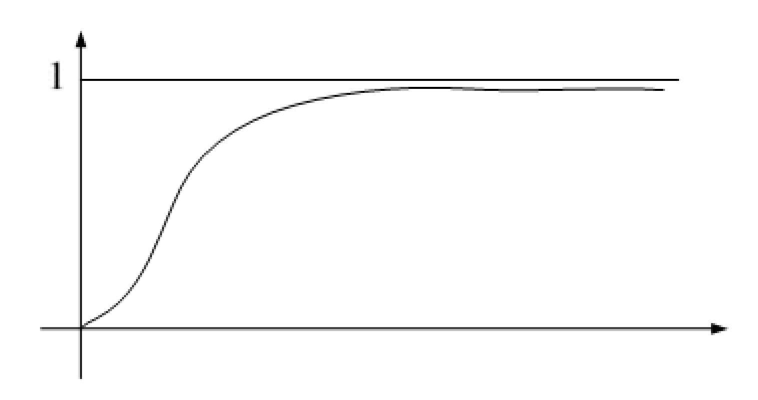
الشكل (1-9) المنحني التكراري المائل إلى اليمين

3 . المنحنيات المائلة إلى اليسار: وهي التي يكون طرفها اليساري طويلاً بالنسبة لمركزها، مثل:



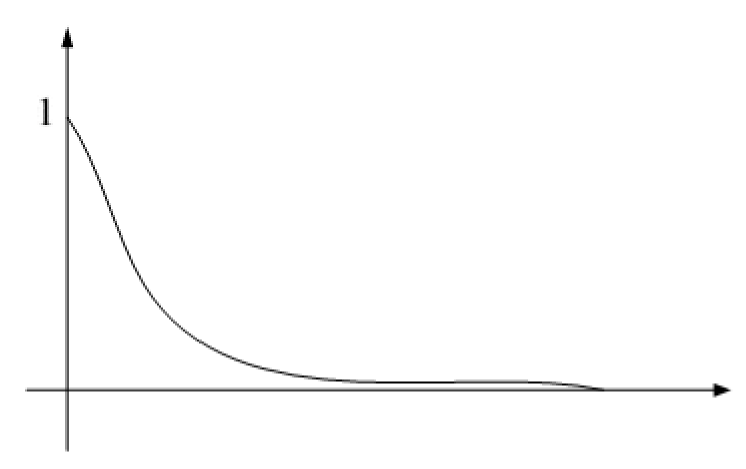
الشكل (1-10): المنحني التكراري المائل إلى اليسار

4 . المنحني التجميعي الصاعد: ويأخذ شكلاً متصاعداً ويبدأ من قيمة الصفر حتى الواحد (أكبر قيمة له) مثل:



الشكل (1-11): المنحني التجميعي الصاعد

5 . المنحنى التجميعي المتنازل: وهو يأخذ شكلاً متنازلاً أو هابطاً ويبدأ من القيمة (1) وينتهي عند القيمة (0) كما يلي:



الشكل (1-12): المنحني التجميعي المتنازل

الفصل الثاني

حساب مقاييس النزعة المركزية والتشتت

أولاً: مقاييس النزعة المركزية(القيم المركزية):

إن مقاييس النزعة المركزية لمتحولX , هي جملة من المؤشرات الإحصائية الكمية التي تعبر عن القيم المركزية التي تتمحور حولها قيم المتحول X .

وإن أهم هذه المقاييس هي :

- المتوسط الحسابي : (للبيانات المفردة والمرتبة والمبوبة).
- الوسيط : وهو يحسب (للبيانات المفردة والمرتبة والمبوبة).
- المنوال : وهو يحسب (للبيانات المرتبة والمبوبة).

ويستفاد من هذه المقاييس عند حسابها في وصف سلوك المتحولات الكمية المدروسة وفي تحديد القيم المركزية لها. ولذلك سنقوم بدراستها وإجراء بعض التطبيقات عليها حسب أنواع البيانات المتوفرة عنها.

2-1: أنواع البيانات المتوفرة:

- البيانات المفردة: وهي البيانات الخام التي نحصل عليها عندما نقوم بجمع المعلومات الإحصائية عن سلوك أي متحول X، وهي تكون على شكل أعداد غير منظمة (عشوائية) لذلك نسميها بيانات مفردة ونرمز لها كما يلي :

$$X: x_1\ x_2\ x_3\ x_4 \ldots\ldots\ldots x_i \ldots\ldots x_n$$

حيث n عدد القيم المتوفرة .

- البيانات المرتبة: وهي البيانات المنظمة تصاعدياً أو تنازلياً في جداول خاصة ، وذلك لأنه في معظم الأحيان نقوم (أو يجب أو نقوم) بترتيب القيم المفردة تصاعدياً (أو تنازلياً) وحساب التكرارات المقابلة لكل من القيم الأساسية فيها, فنحصل على جدول كالتالي :

جدول (2-1) : الشكل العام للبيانات المرتبة حسب القيم الأساسية لـ X.

الرقم المتسلسل i	1	2	3	4	5	6	……	m	المجموع
x_i= قيم x المرتبة	x_1	x_2	x_3	x_4	x_5	x_6	……	x_m	
n_i = التكرارات المطلقة المقابلة	n_1	n_2	n_3	n_4	n_5	n_6	……	n_m	n

حيث أن: n_i عدد التكرارات المقابلة للقيمة x_i وحيث m عدد القيم الأساسية المرتبة.

- البيانات المبوبة: وهي البيانات المصنفة ضمن فئات أو مجالات متلاصقة وغير متقاطعة ، حيث نقوم في بعض الحالات بتبويب أو تصنيف البيانات الإحصائية ضمن فئات أو مجالات جزئية محددة وحساب التكرارات المقابلة لكل فئة أو مجال, فنحصل على جدول كالتالي :

جدول (2-2) : الشكل العام للبيانات المبوبة ضمن فئات أو مجالات حسب قيم X :

الرقم المتسلسل i	1	2	3	4	5	....	m	المجموع
فئات أو مجالات X $[x_i, x_{i+1}[$	$[x_1, x_2[$	$[x_2, x_3[$	$[x_3, x_4[$	$[x_4, x_5[$	$[x_5, x_6[$	....	$[x_m, x_{m+1}[$	
التكرارات المطلقة المقابلة n_i = i	n_1	n_2	n_3	n_4	n_5	....	n_m	n
مراكز الفئات أو المجالات $x'_i = \frac{x_i + x_{i+1}}{2}$	x'_1	x'_2	x'_3	x'_4	x'_5	....	x'_m	

حيث أن: n_i هو عدد التكرارات المقابلة للفئة أو المجال i وأن $n = \sum^m n_i$

وحيث أن: x'_i هي مراكز المجالات الجزئية وتحسب من العلاقة: $x'_i = \frac{x_i + x_{i+1}}{2}$

وحيث أن: m هو عدد المجالات أو الفئات المعتمدة في التبويب.

وبعد هذا الاستعراض السريع والترميز المختصر سنقوم بتعريف وحساب كل من مقاييس النزعة المركزية المذكورة، وذلك حسب أنواع المعلومات المتوفرة.

2-2: المتوسط الحسابي mean :

من المعروف إن المتوسط الحسابي لقيم المتحول X يساوي مجموع تلك القيم مقسوماً على عددها n, أي أن:

$$\text{المتوسط الحسابي} = \frac{\text{مجموع القيم}}{\text{عددها}}$$

وبناء على ذلك وعلى الرموز المستخدمة سابقاً يمكننا التعبير عن المتوسط الحسابي – بعد أن نرمز له بـ $\overline{x}$ – بإحدى العلاقات التالية حسب نوع البيانات المتوفرة .

1-2-2: المتوسط الحسابي لبيانات مفردة :

ويعطي بالعلاقة التالية

$$\overline{x} = \frac{x_1 + x_2 + x_3 + x_4 + \cdots + x_i + \cdots + x_n}{n}$$

والتي يمكن كتابتها بعد استخدام رمز المجموع كما يلي :

$$\overline{\mathrm{x}} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (1-2)$$

مثال (2−1) : لنفترض أننا أجرينا اختباراً لمستوى سكر الدم قبل الإفطار لعينة عشوائية من الطلاب الأصحاء بحجم n=30 طالباً, فوجدنا أن مستويات السكر لديهم كانت كما يلي (ملغ/د.ل).

X :	70	65	85	90	75	85	75	90	100	95
	65	80	85	85	80	90	95	90	95	105
	80	85	70	100	85	95	80	85	70	100

والمطلوب حساب المتوسط الحسابي لهذه المستويات وتقدير متوسط سكر الدم عند الطلاب الأصحاء.

الحل : نقوم بحساب المتوسط الحسابي مباشرة من العلاقة (2-1) السابقة، فنجد أن:

$$\overline{x} = \frac{70+65+85+90+75+\cdots+85+70+100}{30}$$

$$\overline{x} = \frac{\sum^{n} x_i}{n} = \frac{2550}{30} = 85\, mg/dl$$

وهذا يعني أن متوسط مستوى السكر في الدم لدى الشباب الأصحاء يقدر بـ 85 ملغ في كل ديسيلتر من الدم .

2-2-2 : المتوسط الحسابي للبيانات المرتبة :

لحساب هذا المتوسط نعود إلى الجدول (2-1) ونأخذ بعين الاعتبار تكرارات القيم X_i ونضربها بمقدار التكرارات المقابلة لها n_i (وذلك بدلاً من جمعها n_i مرة), فنجد أن العلاقة الرياضية التي تعطينا المتوسط الحسابي للبيانات المرتبة، تأخذ الشكل التالي :

$$\overline{x} = \frac{n_1x_1+n_2x_2+n_3x_3+\cdots\ldots n_mx_m}{n}$$

والتي يمكن كتابتها على الشكل التالي :

$$\overline{\mathrm{x}} = \frac{\sum_{i=1}^{m} n_i\, x_i}{n} \qquad (2-2)$$

مثال(2−2) : لنأخذ معطيات المثال (1.2) حول مستويات السكر في الدم ولنرتبها تصاعديا ونضع التكرارات المقابلة لها فنحصل على الجدول التالي :

جدول (2-3): البيانات المرتبة لمستويات السكر

الرقم المتسلسل i	1	2	3	4	5	6	7	8	9	المجموع
قيم مستويات السكر المرتبة x_i	65	70	75	80	85	90	95	100	105	X
التكرارات المطلقة n_i	2	3	2	4	7	4	4	3	1	30
الجداءات $n_i x_i$	170	210	150	320	595	360	380	300	105	2550

وبعد حساب الجداءات ($n_i x_i$) ووضعها في السطر الأخير ثم حساب مجموعها نجد أن المتوسط الحسابي لهذه البيانات المرتبة يساوي :

$$\overline{x} = \frac{2*65+3*75+2*75+5*80+5*85+5*90+4*95+3*100+1*105}{30}$$

وهو نفس الجواب السابق $$\overline{x} = \frac{\sum n_i x_i}{n} = \frac{2550}{30} = 85\ mg/dl$$

3-2-2 : المتوسط الحسابي للبيانات المبوبة :

لحساب هذا المتوسط نعود إلى الجدول (2-2) ونأخذ بعين الاعتبار التكرارات n_i المقابلة للمجالات أو الفئات ونستبدل كل مجال $[x_i , x_{i+1}]$ بمركزه $\acute{x}_i$ (المحسوب من العلاقة $\acute{x}_i = \frac{x_i + x_{i+1}}{2}$) ثم نقوم بحساب المتوسط الحسابي $\bar{\acute{x}}$ من العلاقة :

$$\overline{\acute{x}} = \frac{n_1 \acute{x}_1 + n_2 \acute{x}_2 + n_3 \acute{x}_3 + n_4 \acute{x}_4 + \cdots .... n_m \acute{x}_m}{n}$$

والتي يمكن كتابتها على الشكل التالي :

$$\overline{\acute{x}} = \frac{\sum_{i=1}^{m} n_i \acute{x}_i}{n} \qquad (3\text{-}2)$$

مثال (3-2) : لنأخذ معطيات المثال (2−1) السابق حول مستويات السكر ولنقم بتبويبها ضمن أربعة مجالات متصلة وغير متقاطعة كما يلي :

جدول (4-2): البيانات المبوبة لمستويات السكر :

الرقم المتسلسل	1	2	3	4	المجموع
المجالات [x_i ، x_{i+1} [	[65,80[	[80,90[	[90,100[	[100,110[	X
التكرارات المقابلة n_i	7	11	8	4	30
مراكز المجالات $\acute{x}_i$	72.5	85	95	105	X
الجداءات $n_i . \acute{x}_i$	507.5	935	760	420	2622.5

وبعد حساب الجداءات $n_i . \acute{x}_i$ ووضعها في السطر الأخير نجد أن :

$$\overline{x} = \frac{7*(72.5)+11*(85)+8*(95)+4*(105)}{30}$$

$$\overline{x} = \frac{\sum n_i \acute{x}_i}{n} = \frac{2622.5}{30} = 87.42\ mg/dl$$

وهذا يختلف قليلا عن المتوسط الحقيقي (85) وذلك بسبب التبويب واستبدال المجالات بمراكزها.

2–2–4 : خواص المتوسط الحسابي :

1– إن قيمة المتوسط الحسابي للبيانات المرتبة تساوي قيمته الحقيقية المحسوبة من البيانات المنفردة (قبل ترتيبها)، ولكن قيمة المتوسط الحسابي للبيانات المبوبة تختلف قليلاً عن القيمة الحقيقية، وذلك لأن العلاقة (2-3) تتضمن مراكز المجالات $\acute{x}_i$ بدلاً من جميع القياسات الحقيقية x_i التي تقع في المجال [x_i ، x_{i+1}] .

2– إذا أضفنا (أو طرحنا) عدداً ثابتاً x_0 من جميع القياسات x_i فإن قيمة المتوسط $\overline{x}$ تزداد (أو تنقص) بمقدار ذلك العدد x_0 وتصبح قيمة المتوسط $\overline{x} \pm x_0$ (البرهان بسيط). ويستفاد من هذه الخاصة في تسهيل الحسابات.

3– إذا ضربنا (أو قسمنا) جميع القياسات x_i بعدد ثابت k فإن قيمة المتوسط الحسابي تتضاعف (او تتناقص) بمقدار k مرة وتصبح قيمة المتوسط $k\,\overline{x}$ (أو $\frac{1}{k}\overline{x}$) (البرهان بسيط) .

4– إن مجموع انحرافات القيم x_i عن متوسطها الحسابي $\overline{x}$ يساوي الصفر : والبرهان على ذلك سهل جداً لأن :

$$\sum_{i=1}^{n}(x_i - \overline{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \overline{x} = n\overline{x} - n\overline{x} = 0$$

اي أن :

$$\sum_{i=1}^{n}(x_i - \overline{x}) = 0 \qquad (4\text{-}2)$$

5– إن مجموع مربعات انحرافات القياسات عن متوسطها الحسابي أصغر من مجموع مربعات انحرافاتها عن أية قيمة أخرى (تقبل بدون برهان) .

6– يمكن حساب قيمة المتوسط الحسابي للبيانات المرتبة أو المبوبة باستخدام التكرارات النسبية $\left(\frac{n_i}{n}\right)$ المقابلة للقيم x_i (أو $\acute{x}_i$) وذلك باستخدام العلاقة المشابهة للعلاقة (2–3) التالية :

$$\overline{x} = \sum_{i=1}^{n}\left(\frac{n_i}{n}\right).x_i \qquad (5\text{-}2)$$

مثال (2–4) : لنأخذ معلومات المثال (2–1) المرتبة في الجدول (2-3) السابق ونحسب التكرارات النسبية المقابلة للقيم المرتبة فنحصل على ما يلي :

جدول (2-5): البيانات المرتبة مع التكرارات النسبيه لمستويات السكر :

x_i القيم المرتبة	65	70	75	80	85	90	95	100	105	المجموع
n_i التكرارات المطلقة	2	3	3	4	7	4	4	3	1	30
التكرارات النسبيه $\left(\frac{n_i}{n}\right)$	$\frac{2}{30}$	$\frac{3}{30}$	$\frac{3}{30}$	$\frac{4}{30}$	$\frac{7}{30}$	$\frac{4}{30}$	$\frac{4}{30}$	$\frac{3}{30}$	$\frac{1}{30}$	1
الجداءات $\left(\frac{n_i}{n}\right).\mathrm{x_i}$	4.333	7	5	10.667	19.833	12	12.667	10	3.5	85

ومن الجدول السابق نجد أن :

$$\overline{x} = \sum \left(\frac{n_i}{n}\right) \mathrm{x_i} = 85$$

وهو نفس الجواب السابق .

7– كما يمكن حساب قيم المتوسط الحسابي للبيانات المرتبة أو المبوبة باستخدام النسب المئوية Pi المقابلة للقيم Xi (أو $\acute{x}_i$) وذلك باستخدام العلاقة التالية :

$$\overline{\mathrm{x}} = \frac{\sum \mathrm{p_i x_i}}{100} \qquad (6\text{-}2)$$

مثال (2–5) : لنأخذ بيانات المثال (2–1) المرتبة في الجدول (2-3) السابق ونحسب النسب المئوية المقابلة للقيم المرتبة فنحصل على الجدول التالي :

جدول (2-6): البيانات المرتبة مع النسب المئوية لمستويات السكر :

x_i القيم المرتبة	65	70	75	80	85	90	95	100	105	المجموع
n_i التكرارات المطلقة	2	3	2	4	7	4	4	3	1	30
P_i % النسب المئوية	6.67	10	6.67	13.33	23.33	13.33	13.33	10	3.33	100
$P_i . X_i$	..	..	..	..	..	..	..	..	..	84.99

ملاحظة : نترك للطالب حساب الجداءات $P_i . X_i$ والتأكد من إن مجموعها يساوي 8500 تقريباً.

ومن الجدول السابق نجد أن $\overline{\mathbf{x}} = \frac{\sum p_i x_i}{100} = \mathbf{84.99} \approx \mathbf{85}$

وهو نفس الجواب السابق (بغض النظر عن الخطأ البسيط بسبب تقريب الأرقام النسبية) .

8- إذا قسمنا القيم المتوفرة عن المتحول X إلى مجموعات جزئية، فإنه يمكننا حساب المتوسط العام $\bar{x}$ من حساب متوسط متوسطات المجموعات الجزئية وبعد وضع التثقيلات المناسبة wi. ويتم حساب حساب المتوسط العام من العلاقة:

$$\bar{\bar{x}} = \frac{\sum w_i \bar{x}_i}{\sum w_i} \qquad (7\text{-}2)$$

حيث أن: $\bar{x}_i$ هو متوسط المجموعة الجزئية i. وأن w_i هي تثقيلات متناسبة مع أحجام أو أثقال المجموعات الجزئية المذكورة.

مثال (2–6): لنقسم مستويات السكر الواردة في المثال (2–1) إلى ثلاث مجموعات ثم نحسب متوسط كل مجموع جزئية كما يلي :

المجموعة الأولى : وتتألف من 5 قياسات هي 75 90 85 65 70

المجموعة الثانية: وتتألف من 10 قياسات هي : 90 85 85 80 65 95 100 90 75 85

المجموعة الثالثة : وتتألف من 15 قياس هي :

90 95 90 95 105 80 85 70 100 85 95 80 85 70 100

ثم نقوم بحساب متوسطات هذه المجموعات الجزئية فنجد أن متوسط المجموعة الأولى يساوي:

$$\overline{x_1} = \frac{70 + 65 + 85 + 90 + 75}{5} = 77$$

وأن متوسط المجموعة الثانية يساوي :

$$\overline{x_2} = \frac{85+75+90+100+95+65+80+85+85+90}{10} = 84$$

وأن متوسط المجموعة الثالثة يساوي :

$$\overline{x_3} = \frac{90 + 95 + 90 + 95 + 105 + 80 + 85 + 70 + 100 + 85 + 95 + 80 + 85 + 70 + 100}{15} = 88.333$$

وبعد معرفتنا لهذه المتوسطات لا يجوز حساب المتوسط العام $\bar{\bar{x}}$ منها بأخذ مجموعها وتقسيمه على /3/, لأن المتوسطات الجزئية $\overline{x_1}$ ، $\overline{x_2}$ ، $\overline{x_3}$ مأخوذة من مجموعات جزئية ذات حجوم مختلفة هي 5 و 10 و 15 . لذلك يقتضي الأمر عند حساب المتوسط العام لهذه المتوسطات أن نثقل كل منها بحجم المجموعة التي يقابلها, فنضع التثقيلات والمتوسطات كما يلي :

$w_1 = 5$ $\quad$ $w_2 = 10$ $\quad$ $w_3 = 15$

$\overline{x_1} = 77$ $\quad$ $\overline{x_2} = 84$ $\quad$ $\overline{x_3} = 88.333$

ثم نقوم بحساب المتوسط العام من العلاقة :

$$\bar{\bar{x}} = \frac{w_1\overline{x_1} + w_2\overline{x_2} + w_3\overline{x_3}}{w_1 + w_2 + w_3}$$

$$\bar{\bar{x}} = \frac{5*(77)+10*(84)+15*(88.33)}{5+10+15} = \frac{2550}{30} = 85$$

ملاحظة : يمكننا أن نضع أية تثقيلات متناسبة مع حجوم الفئات مثل :

$W_2 = 1$, $\overline{x_1} = 77$ $W_2 = 2$, $\overline{x_2} = 84$ $W_3 = 3$, $\overline{x_3} = 88.333$

وعندها نقوم بحساب المتوسط العام من العلاقة :

$$\bar{\bar{x}} = \frac{1*(77)+2*(84)+3*(88.333)}{1+2+3} = \frac{510}{6} = 85$$

وهو نفس الجواب (أعد المثال بوضع التثقيلات 6 ، 4 ، 2)

9- إن المتوسط الحسابي يمثل بشكل جيد مركز القياسات إذا كانت متجانسة وتتوزع على الجانبين توزيعاً متناظراً أو شبه متناظر، ولكنه يتأثر بالقيم المتطرفة أو الشاذة وخاصة بالقيم الكبيرة حيث ينحاز باتجاه تلك القيم ويفقد قيمته الإحصائية .

مثال (2-7) : لنأخذ القيم التالية لدخول /5/ أخوة (ألف ليرة) 300 ، 17 ، 12 ، 15 ، 20, فنجد أن متوسط الدخل لهؤلاء الأخوة يساوي :

$$\bar{x} = \frac{20+15+12+17+300}{5} = 72.8 \quad \text{(الف ليرة)}$$

وهي قيمة متحيزة لأنها تميل نحو القيمة الكبيرة /300/ لأن القياسات غير متجانسة , وفي مثل هذه الحالات يجب العمل على عزل القيم الشاذة عن القياسات المتجانسة قبل حساب المتوسط الحسابي . وهنا إذا قمنا بعزل القيمة /300/ (التابعة للأخ الغني) واستثنائها من الحساب نجد أن متوسط الدخل يساوي :

$$\overline{x_1} = \frac{\mathbf{20+15+12+17}}{\mathbf{4}} = \frac{\mathbf{64}}{\mathbf{4}} = 16 \text{ (الف س . ل)}$$

وهي قيمة مقبولة وذات معنى اقتصادي لأن القياسات أصبحت متجانسة .

10- إن مجموع مربعات انحرافات القيم Xi عن المتوسط الحسابي $\bar{x}$ هو أصغر من مجموع مربعات انحرافات القيم Xi عن أي قيمة أخرى تختلف عن $\bar{x}$ (تقبل بدون برهان)

11– إن نسبة أية خاصة نوعية في المجتمع أو العينة هي حالة خاصة من المتوسط الحسابي لمتحول X, يأخذ القيمة (1) إذا كان العنصر يتميز بتلك الخاصة، ويأخذ القيمة (0) إذا كان العنصر لا يتميز بتلك الخاصة .

فإذا كان عدد العناصر المدروسة n عنصراً ، وكان عدد العناصر التي تتميز بالخاصة النوعية يساوي k عنصراً من أصل n ، فإن نسبة تلك الخاصة تساوي : $r = \frac{K}{n}$, ويمكن حسابها من المتوسط كما يلي :

$$\overline{\mathrm{x}} = \frac{\sum x_i}{n} = \frac{1+0+0+1+0+1+\cdots+1+0}{n} = \frac{k}{n} = r \qquad (8\text{–}2)$$

12– إن مجالات تطبيق المتوسط الحسابي كثيرة جداً، ويعتمد عليه في كثير من الحسابات والدراسات نظراً لبساطته ولسهولة حسابه، وهو يطبق على البيانات المطلقة المكانية . أما إذا كانت البيانات زمانية أو نسبية فإننا نطبق متوسطات أخرى(كالتوافقي أو الهندسي) .

$(3-2)$: الوسيط Median :

تعريف الوسيط : هو القيمة العددية من قيم المتحول X التي تقع في وسط القيم المرتبة أو المبوبة (وهو لا يحسب إلا للبيانات المرتبة أو المبوبة) .

وبتعبير آخر : الوسيط هو القيمة التي تقسم السلسلة المرتبة أو المبوبة لقيم X إلى قسمين متساويين , بحيث يكون عدد القيم التي على يسارها مساوياً لعدد القيم على يمينها . وهو لا يأخذ بعين الاعتبار القيم نفسها, لذلك فهو لايتأثر بالقيم المتطرفة .

مثال(2–7): لنأخذ جملة الأعداد المرتبة لدرجات 13 طالباً والمعروضة في الجدول التالي:

جدول(2–7): درجات الطلاب المرتبة

رقم التسلسل i	1	2	3	4	5	6	7	8	9	10	11	12	13
القيم المرتبة x_i	12	14	15	17	19	20	30	40	50	60	70	80	90

نلاحظ أن عدد القيم فردي n = 13 وأن القيمة التي تقع في وسط القيم المرتبة هي قيمة الوسيط $(\mathrm{x}_7 = 30)$, أي أنها القيمة السابعة , أي هي التي يكون ترتيبها المتسلسل:

$$\frac{\mathrm{n+1}}{2} = \frac{13+1}{2} = 7$$

ونكتب ذلك كما يلي:

$$\mathrm{M_e} = \mathrm{X}_{\frac{n+1}{2}} = \mathrm{x}_7 = 30$$

مثال(2–8): لنأخذ جملة الأعداد المرتبة لدرجات 8 طلاب والمبينة في الجدول التالي:

جدول(2–8): درجات الطلاب المرتبة

الرقم المتسلسل i	1	2	3	4	5	6	7	8

القيم المرتبة x_i	12	15	17	30	70	80	90	100

وهنا نلاحظ أن عدد القيم زوجي n = 8 , وهنا يكون لدينا قيمتان تقعان في وسط القيم المرتبة وهما:

$x_4 = 30$ و $x_5 = 70$ ، وهنا نجد أنه من الطبيعي أن نعتبر قيمة الوسيط هي متوسط القيمتين اللتين تقعان في منتصف السلسلة المرتبة . أي أنها التي تساوي :

$$M_e = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2} = \frac{x_4 + x_5}{2} = \frac{30+70}{2} = 50$$

2–3–1: الوسيط للبيانات المفردة والمرتبة: بناء على ما تقدم يمكننا حساب قيمة ذلك الوسيط وفي كلتا الحالتين من العلاقة :

$$M_e = \begin{cases} x_{\frac{n+1}{2}} & \text{إذا كان n فردياً} \\ \dfrac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2} & \text{إذا كان n زوجياً} \end{cases} \qquad (9\text{–}2)$$

2–3–2: الوسيط لمعلومات مرتبة مع التكرارات :

لنأخذ سلسلة البيانات المرتبة لدرجات قسم العملي لـ 25 طالباً مع تكراراتها المطلقة والتجميعية المتصاعدة كما في الجدول التالي :

الجدول(2–9):البيانات وتكراراتها التجميعية المتصاعدة.

الرقم المتسلسل i	1	2	3	4	5	6	7	8	المجموع
القيم الأساسية المرتبة x_i	14	17	18	19	25	30	35	40	
التكرارات المطلقة n_i	3	4	3	4	5	3	2	1	25
التكرارات التجميعية المتصاعدة k_i	3	7	10	14	19	22	24	25	×

نلاحظ أن العدد الكلي للقيم = 25 قيمة وأن نصفها $\frac{n}{2} = 12.5$.

وحتى نحدد قيمة الوسيط M_e نتابع الأرقام من سطر التكرارات التجميعية المتصاعدة k_i , حتى نصل إلى عدد أكبر مباشرة من $(\frac{n}{2} = 12.5)$ فنجد أنه العدد $k_4 = 14$, وهذا يعني أن قيمة الوسيط هي قيمة X المقابلة لهذا التكرار التجميعي $k_4 = 14$ وهي $x_4 = 19$ ونكتب ذلك كما يلي:

$$M_e = x_{k_i} \;:\; \frac{n}{2} \leq k_i < \frac{n}{2} + n_i \qquad (10\text{–}2)$$

2–3–3: الوسيط للبيانات المبوبة :

لنأخذ درجات 124 طالباً مبوبة مع تكراراتها التجميعية المتصاعدة كما في الجدول التالي:

الجدول(2–10): البيانات المبوبة مع تكراراتها التجميعية المتصاعدة.

الرقم المتسلسل i	1	2	3	4	5	6	المجموع
مجالات التبويب $[x_i , x_{i+1}[$	50–55	55–60	60–65	65–70	70–75	75–80	×

التكرارات المطلقة n_i	20	40	30	20	10	4	124
التكرارات التجميعية المتصاعدة k_i	20	60	90	110	120	124	×

وهنا نلاحظ أن عدد القيم n = 124 ، وأن نصفها هو $\frac{n}{2} = 62$ ، أي أن رقم (ترتيب) الوسيط هو حوالي 62 . لذلك نتابع الأرقام في سطر التكرارات التجميعية k_i حتى نصل ونتجاوز مباشرة العدد 62 فنجد أنه k_3 = 90 , أي أن الوسيط يقع في المجال الثالث وإن قيمته M_e محصورة بين حديه, أي أن: $60 \leq M_e \leq 65$ ، ويسمى هذا المجال مجال الوسيط أو المجال الوسيطي.

ولتحديد قيمة الوسيط M_e للبيانات المبوبة السابقة ضمن ذلك المجال, نفترض أن y قطعة من المجال الوسيطي لتساعدنا في حساب قيمة الوسيط, بحيث تحقق العلاقة التالية:

$$M_e = 60 + y = x_m + y$$

حيث x_m هو الحد الأدنى للمجال الوسيطي .

وعندما نقوم بحساب قيمة y نكون قد حسبنا قيمة الوسيط M_e ضمن المجال الوسيطي .

ولحساب قيمة y نجري التناسب التالي :

بما أن قيمة الوسيط M_e ضمن المجال تقابل $\frac{n}{2}$ من التكرارات .

فإن القطعة $y = M_e - X_m$ من المجال يجب أن تقابل $\left(\frac{n}{2} - k_{m-1}\right)$ من التكرارات .

وبما أن طول المجال الوسيطي $d_m = x_{m+1} - x_m$ يقابل فقط n_m من التكرارات .

فإن يكون لدينا التناسب التالي :

$$\frac{y}{d_m} = \frac{\frac{n}{2} - k_{m-1}}{n_m}$$

ومنها نجد أن :

$$y = d_m \frac{\frac{n}{2} - k_{m-1}}{n_m}$$

وبالعودة إلى العلاقة المفروضة نحصل على أن :

$$M_e = x_m + d_m \frac{\frac{n}{2} - k_{m-1}}{n_m} \qquad (11\text{-}2)$$

واعتماداً على هذه العلاقة يمكننا إيجاد الوسيط لسلسلة درجات الطلاب المفروضة في الجدول السابق مع تكراراتها المطلقة والتجميعية، فنجد أن :

$$M_e = 60 + (65 - 60)\,\frac{62 - 60}{30} = 60.333$$

وهي قيمة واقعة في المجال الوسيطي وقريبة جداً من طرفه الأيسر .

(4 – 2) : المنوال Mode :

تعريف المنوال: هو القيمة من قيم X التي تقابل التكرار الأكبر, أي هي القيمة الأكثر تكراراً، ولذلك تسمى (الموضة) , وهو لا يعرف إلا على البيانات المرتبة أو المبوبة.

2-4-1: المنوال للبيانات المرتبة:

لحساب المنوال للبيانات المرتبة نأخذ المثال التالي:

مثال (2-9): لنأخذ درجات الطلاب في الامتحان العملي (لـ 50 طالباً) والمبينة في الجدول التالي :

الجدول(2-11): درجات الطلاب المرتبة مع تكراراتها.

الرقم المتسلسل i	1	2	3	4	5	6	المجموع
قيم x المرتبة : x_i	2	4	8	16	32	40	×
التكرارات المطلقة	3	5	10	15	10	7	50

نلاحظ أن القيمة التي تقابل التكرار الأكبر (15) هي قيمة X الرابعة أي أن المنوال هو:

$$M_o = x_4 = 16$$

أي أن الدرجة 16 هي قيمة المنوال لأنها الأكثر تكراراً بين درجات الطلاب في العملي .

2-4-2: المنوال للبيانات المبوبة :

مثال(2-10): لنأخذ البيانات التالية عن درجات 220 طالباً، والمبوبة كما في الجدول التالي :

الجدول(2-12): درجات الطلاب المبوبة مع تكراراتها.

الرقم المتسلسل i	1	2	3	4	5	6	7	المجموع
مجالات التبويب $[x_i , x_{i+1}[$	$[20-30[$	$[30-40[$	$[40-50[$	$[50-60[$	$[60-70[$	$[70-80[$	$[80-90[$	×
التكرارات المطلقة n_i	32	40	45	60	25	15	3	220

نلاحظ أن التكرار الأكبر هو (60) وهو يقابل المجال الرابع المحدد بـ $]60 - 50[$ ، ويسمى هذا المجال بالمجال المنوالي $]x_{m+1} , x_m[$.

أي أن المنوال M_o يقع في هذا المجال وأن قيمته تحقق العلاقة : $50 \leq M_o \leq 60$

ولحساب قيمة M_o ضمن المجال المنوالي $]x_{m+1} , x_m[$ نلاحظ أن طوله يساوي:

d_m = 60 – 50 = 10 ، وإن التكرار الذي يسبقه: n_{m-1} = 45 ، وإن التكرار الذي يليه :

n_{m+1} = 25 ، وبذلك نجد أنه يمكننا حساب قيمة المنوال M_o من العلاقة : (بدون برهان)

$$M_o = x_m + d_m \frac{n_m - n_{m-1}}{(n_m - n_{m-1}) + (n_m - n_{m+1})} \qquad (12_2)$$

واعتماداً على معطيات الجدول السابق يمكننا حساب قيمة المنوال كما يلي:

$$M_o = 50 + 10 \frac{60- 45}{(60- 45) + (60- 25)} = 50 + 10 \frac{15}{50}$$

$$M_o = 50 + 3 = 53$$

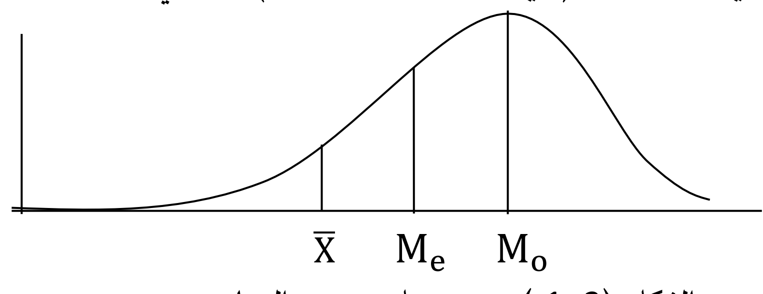
ملاحظة : يمكن أن يكون للمعلومات المرتبة أو المبوبة أكثر من منوال .

(2-5): تطبيقات مشتركة للمتوسط الحسابي والوسيط والمنوال :

إن قيم المتوسط الحسابي والوسيط والمنوال كقيم مركزية لنفس المتحولX ، قد تكون مختلفة أو متقاربة أو متساوية . وهنا نميز الحالات الأساسية التالية :

1- إذا كانت العلاقة بينهم كما يلي : $\bar{x} < M_e < M_o$

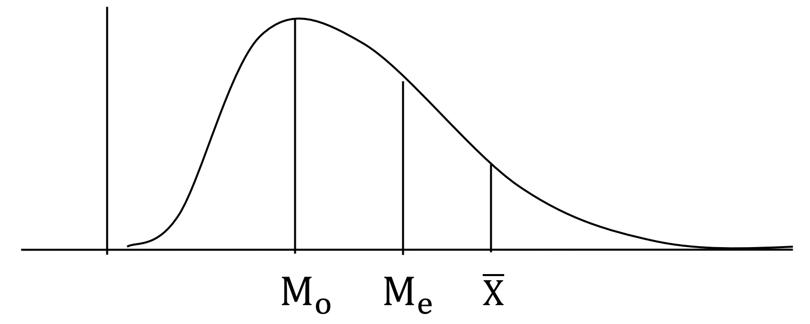
فإن التوزيع التكراري يكون ملتوياً نحو اليسار (أي أنه يمتد نحو اليسار) كما يلي :



الشكل (2-1):منحنى ملتوي نحو اليسار

2- إذا كانت العلاقة بينهم كما يلي : $M_o < M_e < \bar{x}$

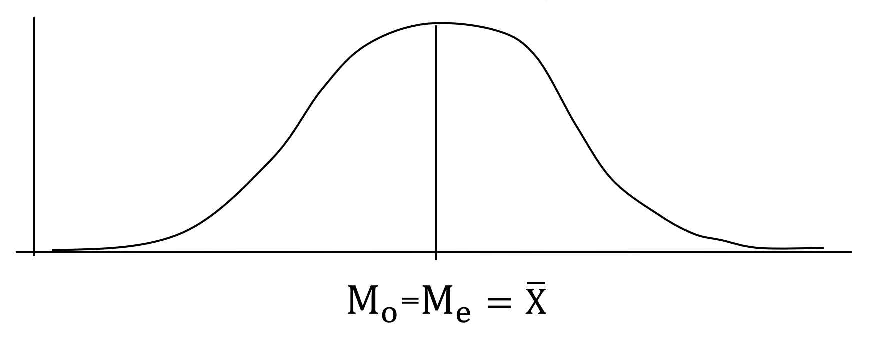
فإن التوزيع التكراري يكون ملتوياً نحو اليمين (أي أنه يمتد نحو اليمين) كما يلي :



الشكل (2-2): منحنى ملتوي نحو اليمين

3- إذا كانت العلاقة بينهم كما يلي : $\bar{x} = M_o = M_e$

فإن التوزيع التكراري يكون متناظراً تماماً كما يلي :



الشكل (2–3): منحنى متناظر

ملاحظة هامة : إن قيمة الوسيط M_e تقع دائماً بين المتوسط الحسابي $\bar{X}$ والمنوال M_o أو تنطبق عليهما.

ثانياً: مقاييس التشتت:

2–5: تمهيد:

لتوضيح فكرة التشتت ولبيان ضرورة إيجاد مقاييس كمية لحساب مقادير التشتت لكل متحول عشوائي X، نأخذ درجات الكوليسترول الحميد MDL في الدم (mg/dl) لمجموعتين من الذكور والإناث، ولنفترض أنها كانت كما يلي :

X:	40	41	42	43	44	45	46	47	48	49	50
Y:	30	33	36	39	42	45	48	51	54	57	60

وإذا قمنا بحساب المتوسط الحسابي لكل منهما لوجدنا أن :

$$\bar{x} = \frac{\mathbf{40+41+42+43+44+45+\cdots\ \ 48+49+50}}{\mathbf{11}} = 45\ mg/dl$$

$$\bar{y} = \frac{\mathbf{30+33+36++39+42+45+\cdots\ \ 54+57+60}}{\mathbf{11}} = 45\ mg/dl$$

وبالمقارنة نجد أن هذين المتوسطين متساويان أي أن :

$$\bar{x} = \bar{y} = 45$$

ولكن القراءة السريعة لبيانات هاتين المجموعتين تظهر لنا بوضوح أنهما تختلفان عن بعضهما البعض .

فكيف نفرق أو نميز بين بيانات هاتين المجموعتين علماً بأن المتوسطين متساويان ؟

وللجواب على ذلك نلاحظ :

1– إن أصغر قيمة للمجموعة X هي $X_I = 40$ وإن أكبر قيمة فيها $X_{II} = 50$, بينما نجد أن أصغر قيمة للمجموعة Y هي $Y_I = 30$ وإن أكبر قيمة فيها $Y_{II} = 60$.

2– أن قيم X تتغير في المجال [50 , 40] ، بينما قيم Y تتغير في المجال [60 , 30] وهو مجال أكبر من المجال الأول. وهذا يعني أن تشتت قيم المجموعة Y أكبر من تشتت قيم المجموعة X. وبناءً على ما تقدم تم تعريف عدة مقاييس لحساب التشتت هي:

(2–6): المدى Range :

تعريف المدى : هو عبارة عن الفرق بين أكبر قيمة وأصغر قيمة للمتحول المدروس X ويعطى بالعلاقة التالية :

$$R = x_{max} - x_{min} \qquad (2–13)$$

ومن عيوب هذا المقياس ما يلي :

- إنه لا يأخذ بعين الاعتبار إلا القيمتين الكبرى X_{max} والصغرى X_{min} ويهمل تأثير القيم الأخرى على التشتت وعلى قيمته العددية.
- إن شكله العام (2–13) لا يفرق بين قيمتين متساويتين للمدى R لظاهرتين أو لمتحولين مختلفتين . فإذا كان الفرق بين درجات حرارة الإنسان عن معدلها العام يساوي 4 درجات فهذا أمر خطير قد يؤدي إلى الوفاة, ولكن الفرق في درجة حرارة فرن الحديد بمقدار 4 درجات لا يعتبر ذا أهمية.

ولذلك تم تعريف المؤشر النسبي للمدى المنسوب إلى المتوسط $\bar{x}$, والذي يحسب من العلاقة :

$$\boldsymbol{r} = \frac{\boldsymbol{R}}{\bar{\boldsymbol{x}}}\ 100 = \frac{x_{max} - x_{min}}{\bar{x}}\ 100\ \% \qquad (14\text{–}2)$$

وإن هاتين العلاقتين تستخدمان لقياس المدى من المعلومات المفردة أو المرتبة أو المبوبة , ويفضل عند استخدامه الاعتماد على شكله النسبي المعرف بالعلاقة (2–14).

مثال(2–11) : لنأخذ المعلومات السابقة عن مستويات الكوليسترول الحميد HDL ونحسب المدى العام والنسبي لهما فنجد أن :

$R_x = 50 - 40 = 10$ $\qquad r_x = \frac{\mathbf{10}}{\mathbf{45}}\ 100 = 22.22\ \%$

$R_y = 60 - 30 = 30$ $\qquad r_y = \frac{\mathbf{30}}{\mathbf{45}}\ 100 = 66.67\ \%$

2–7 : متوسط الانحرافات بالقيمة المطلقة:

حتى نأخذ تأثير تشتت جميع القيم x_i عن المتوسط $\bar{x}$, يمكن أن نحسب انحرافات تلك القيم x_i عن المتوسط الحسابي $\bar{x}$, ولكن عندما نأخذ مجموع هذه الانحرافات عن $\bar{x}$ بشكلها الجبري، نجد أنه يساوي الصفر في جميع الحالات.لأنه قد برهنا سابقاً على أن:

$$\sum(\mathrm{x_i} - \bar{\mathrm{x}}) = \sum \mathrm{x_i} - \sum \bar{\mathrm{x}} = n\bar{x} - n(\bar{x}) = 0$$

لذلك يمكن أن نأخذ تلك الانحرافات عن المتوسط بالقيمة المطلقة، ثم نأخذ مجموعها ونقسمه على عددها (n) . وبناءً على ذلك نعرف متوسط الانحرافات المطلقة بالعلاقة:

$$\Delta = \frac{\sum|\mathrm{x_i} - \bar{\mathrm{x}}|}{\mathrm{n}} \qquad (15\text{–}2)$$

وهناك علاقة أخرى لتعريف هذا المؤشر يأخذ مجموع الانحرافات عن الوسيط M_e ثم تقسيمها على عددها (n) وهي:

$$\grave{\Delta} = \frac{\sum|\mathrm{x_i} - \mathrm{M_e}|}{\mathrm{n}} \qquad (16\text{–}2)$$

وإن الشكل النسبي لهذين المؤشرين يعرف العلاقتين:

$$\delta = \frac{\Delta}{\bar{x}} 100 \qquad \acute{\delta} = \frac{\acute{\Delta}}{M_e} 100 \qquad (17-2)$$

ومع إن هذا المؤشر يتميز بعدة خواص جيدة ولكن وجود القيمة المطلقة في تعريفه الرياضي جعله غير ملائم لعمليات التحليل الرياضي .

لذلك تم تعريف مؤشرين جديدين يعوضان عنه وهما: التباين والانحراف المعياري، واللذين يعتمدان على حساب مجموع مربعات الانحرافات عن متوسطها الحسابي ثم تقسيمه على عددها (n) .

(2-8) : التباين والانحراف المعياري ***Variance and standard deviation***

يعتبر التباين من أهم مقاييس التشتت، ويرمز له بالرمز σ^2 **(ويلفظ سيجما)**, وهو يعرف رياضياً حسب نوع المعلومات المتوفرة لأي متحول x بالعلاقات التالية:

1- للمعلومات المفردة : $X : x_1 \ x_2 \ x_3 \ .... \ x_i \ ..... \ x_n$ بالعلاقة:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} \qquad (18-2)$$

ويمكن استخلاص علاقة أخرى منها لتسهيل الحسابات وهي:

$$\sigma^2 = \frac{\sum x_i^2}{n} - (\bar{x})^2$$

2- للمعلومات المرتبة بالعلاقة:

$$\sigma^2 = \frac{\sum_{i=1}^{m} n_i(x_i - \bar{x})^2}{n} \qquad (19-2)$$

وهناك علاقة أخرى مشتقة منها لتسهيل الحسابات هي:

$$\sigma^2 = \frac{\sum n_i x_i^2}{n} - \bar{x}^2$$

3- للمعلومات المبوبة بالعلاقة :

$$\sigma^2 = \frac{\sum_{i=1}^{m} n_i\left(x_i' - \bar{x}'\right)^2}{n} \qquad (20-2)$$

وهناك علاقة مشتقة منها لتسهيل الحسابات وهي:

$$\sigma^2 = \frac{\sum n_i x_i'^2}{n} - (\bar{x}')^2$$

ومن العلاقة (2-18) يتضح لنا بأنه يمكننا أن نعرف التباين كما يلي:

التباين : هو متوسط مربعات انحرافات القيم x_i عن متوسطها الحسابي $\bar{x}$.

أما الانحراف المعياري للمتحول X فيعرف بأنه الجذر التربيعي الموجب لقيمة التباين σ^2 ويحسب من العلاقة:

$$\sigma = +\sqrt{\sigma^2} \qquad \text{(في جميع الحالات)} \qquad (21-2)$$

مثال(2–12): احسب المتوسط الحسابي والتباين والانحراف المعياري لمستوى اللمفاويات في الدم لعينة مؤلفة من *N=12* طالباً من طلاب الجامعة ، والتي كانت كما يلي (*mg/dL*):

X: 25 , 27 , 28 , 29 , 30 , 24 , 35 , 38 , 40 , 26 , 28 ,30

الحل : إن المتوسط الحسابي لهذه البيانات يساوي:

$$\bar{x} = \frac{25+27+28+29+30+24+35+38+40+26+28+30}{12} = 30$$

ولحساب التباين يجب علينا أولاً حساب الانحرافات $(x_i - \bar{x})$ ثم تربيعها ثم أخذ مجموعها وتقسيمه على عددها, ولتسهيل هذه العمليات نعد الجدول التالي:

الجدول(2–13): بيانات المثال مع بعض الحسابات اللازمة.

I	1	2	3	4	5	6	7	8	9	10	11	12	المجموع
X_i	25	27	28	29	30	24	35	38	40	26	28	30	
$x_i - \bar{x}$	−5	−3	−2	−1	0	+6	+5	+8	+10	−4	−2	0	0
$(x_x - \bar{x})^2$	25	9	4	1	0	36	25	64	100	16	4	0	284

ومن هذا الجدول نجد أن مجموع مربعات الانحرافات يساوي 284 , وبناءً على ذلك نحسب التباين من العلاقة :

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \frac{284}{12} = 23.667$$

أما الانحراف المعياري فنجد أنه يساوي

$$\sigma = \sqrt{\sigma^2} = \sqrt{23.667} = 4.865$$

مثال(2-13): احسب المتوسط الحسابي والتباين والإنحراف المعياري للبيانات المرتبة التالية عن مستويات اللمفاويات x_i لدى 60 طالباً (mg/dL) التالية:

الجدول(2–14): بيانات الطلاب المرتبة مع بعض الحسابات اللازمة.

I	1	2	3	4	5	6	7	المجموع
x_i	20	23	26	29	32	35	38	
n_i	4	5	10	15	12	8	6	60
$n_i x_i$	80	115	260	435	384	280	228	1782
$(x_x - \bar{x})$	−9.7	−6.7	−3.7	−0.7	2.3	5.3	8.3	
$(x_x - \bar{x})^2$	94.09	44.89	14.69	0.49	5.29	28.09	68.89	
$n_i(x_x - \bar{x})^2$	376.36	224.45	136.9	7.35	63.48	224.72	413.34	1446.6

الحل : لحساب المتوسط الحسابي والتباين نضيف إلى الجدول السابق بعض السطور لحساب الكميات التي نحتاجها وهي:

$$n_i\,(x_i - \bar{x})^2 \quad \text{ثم} \quad (x - x_i)^2 \quad \text{و} \quad (x_i - \bar{x}) \quad \text{و} \quad n_i\,x_i$$

وبذلك نجد أن قيمة المتوسط الحسابي تساوي (انظر مجموع $n_i\,x_i$ في الجدول):

$$\bar{\mathbf{x}} = \frac{\sum_{i=1}^{7} n_i x_i}{n} = \frac{1782}{60} = 29.7\ mg/dl$$

وبعد حساب كل من $(x_i - \bar{x})$ و $(x_i - \bar{x})^2$ و $n_i\,(x_i - \bar{x})^2$ ووضعها في الأسطر الملحقة بالجدول السابق ، نأخذ مجموع السطر الأخير فنحصل على أن:

$$\sum n_i (x_i - \bar{x})^2 = 1446.6$$

وبذلك يمكننا حساب قيمة التباين من العلاقة:

$$\sigma^2 = \frac{\sum_{i=1}^{7} n_i (x_i - \bar{x})^2}{n} = \frac{1446.6}{60} = 24.11$$

ثم نقوم بحساب قيمة الانحراف المعياري σ من العلاقة:

$$\sigma = \sqrt{\sigma^2} = \sqrt{24.11} = 4.91$$

<u>2–8–1: بعض خواص التباين والانحراف المعياري :</u>

1– إذا جمعنا (أو طرحنا) من جميع القيم X_i عدداً ثابتاً X_0 فإن قيمة التباين لا تتغير: لأن المتوسط سيزداد بمقدار X_0 وبذلك فإن الانحرافات عن المتوسط الجديد لا تتغير. وبالتالي فإن قيمة التباين والانحراف المعياري لا تتغير.

2– إذا ضربنا (أو قسمنا) جميع القيم X_i بعدد ثابت K فإن قيمة التباين σ^2 تتضاعف بمقدار K^2 مرة (أو تتناقص بمقدار $\frac{1}{k^2}$) وتصبح قيمة التباين $k^2\ \sigma^2$ (أو $\frac{1}{k^2}\ \sigma^2$)، أما قيمة الانحراف المعياري σ فتتضاعف بمقدار K مرة وتصبح $k.\ \sigma$ (تتناقص بمقدار $\frac{1}{K}$ مرة).

3– يبرهن رياضياً أن قيمة التباين هي أصغر من قيمة متوسط مربعات انحرافات القيم X_i عن أية قيمة أخرى $\bar{y}$ تختلف عن المتوسط الحسابي $\bar{x}$ (بدون برهان).

4– إذا جزّأنا البيانات المتوفرة (المفردة أو المرتبة أو المبوبة) إلى مجموعات جزئية , فإنه يمكننا حساب التباين داخل كل مجموعة K والتباين بين المجموعات نفسها, وبذلك نحصل على عدة تباينات هي:

أ. التباينات داخل كل مجموعة K(عدد عناصرها n_k) ويعطى بالعلاقة:

$$\sigma_k^2 = \frac{\sum(x_i - \bar{x}_k)^2}{n_k} \qquad (2-22)$$

ب. متوسط التباينات الداخلية المثقل بعدد عناصر المجموعاتn_k :

$$\bar{\sigma}^2 = \frac{\sum n_k \,.\, \sigma_k^2}{\sum n_k} \qquad (2-23)$$

ج. التباين بين المجموعات: وهو يعكس انحرافات متوسطات المجموعات $\bar{x}_k$ عن المتوسط الحسابي العام $\bar{x}$ ويحسب من العلاقة:

$$\delta^2 = \frac{\sum n_k\,(x_k - \bar{x})^2}{\sum n_k} \qquad (2-24)$$

د. التباين الكلي (الأصلي قبل التجزئة): وهو يساوي :

$$\sigma^2 = \frac{\sum^n (x_i - \bar{x})}{n} \qquad (2-25)$$

هـ. ويمكن البرهان على أن التباين الكليσ^2 يساوي مجموع التباينين (متوسط التباينات الداخلية $\bar{\sigma}^2$ والتباين بين المجموعات δ^2) أي أن:

$$\sigma^2 = \bar{\sigma}^2 + \delta^2 \qquad (2-26)$$

وسيكون لهذه العلاقات أهمية خاصة في تحليل التباين لاحقاً

مثال(2-14): لنفترض أنه لدينا القياسات المفردة والمرتبة التالية والمجمعة ضمن ثلاث مجموعات بحجوم $n_1 = 3$ $n_2 = 4$ $n_3 = 5$ كما يلي:

X: (2,4,6) , (8,10,12,14) , (16,18,20,22,24)

والمطلوب : إيجاد التباينات المختلفة داخل المجموعات وبين المجموعات ثم حساب التباين الكلي والتأكد من صحة العلاقة (2-26) .

الحل : نقوم أولاً بحساب المتوسطات للمجموعات، فنجد أنها تساوي:

$$\bar{x}_1 = 4 \qquad \bar{x}_2 = 11 \qquad \bar{x}_3 = 20$$

أما المتوسط العام لجميع هذه القياسات فيساوي : $\bar{x} = 13$

ولحساب التباينات الداخلية ضمن كل مجموعة نجد من العلاقة (2–22) أن:

$$\sigma_1^2 = \frac{\sum(x_i - 4)^2}{3} = 2.667$$

$$\sigma_2^2 = \frac{\sum(x_i - 11)^2}{4} = 4$$

$$\sigma_3^2 = \frac{\sum(x_i - 20)^2}{5} = 8$$

وبذلك نجد أن متوسط هذه التباينات يساوي:

$$\overline{\sigma}^2 = \frac{\sum \boldsymbol{n_i}\, \sigma_{\boldsymbol{i}}^{\boldsymbol{2}}}{\sum \boldsymbol{n_i}} = \frac{\mathbf{3*(2.667)+4*4+5*8}}{\mathbf{3+4+5}} = 5.6667$$

ثم نقوم بحساب التباين بين المجموعات δ^2 من العلاقة (2–24)، فنجد أن:

$$\delta^2 = \frac{\sum_{k=1}^{3} n_k(\bar{x}_k - 13)^2}{\sum n_k} = \frac{3*(4-13)^2 + 4*(11-13)^2 + 5*(20-13)^2}{12} = 42$$

ثم نقوم بحساب التباين الكلي σ^2 لهذه المعلومات (قبل التجزئة)، فنجد أنه يساوي:

$$\sigma^2 = \frac{\sum_{\mathbf{i=1}}^{\mathbf{12}} (\mathbf{x_i} - \bar{\mathbf{x}})^{\mathbf{2}}}{\mathbf{12}} = \frac{\sum_{\mathbf{i=1}}^{\mathbf{12}} (\mathbf{x_i} - \mathbf{13})^{\mathbf{2}}}{\mathbf{12}} = 47.6667$$

وللتأكد من صحة العلاقة (2–26) نلاحظ أن:

$$\overline{\sigma}^{\mathbf{2}} + \boldsymbol{\delta}^{\mathbf{2}} = 5.6667 + 42 = 47.6667$$

أي أن :

$$\overline{\sigma}^2 + \delta^2 = \sigma^2$$

2–8–2 : تطبيقات الانحراف المعياري: (S D)

يستخدم الانحراف المعياري في جميع الدراسات والأبحاث, فهو يدخل في تعريف معامل الإختلاف, ويعتمد عليه في إنشاء مجالات الثقة المختلفة, ويدخل في إختبارات الفرضيات وفي إتخاذ القرارت ..الخ.

(2–8–2–1)– معامل الاختلاف (CV) Coefficient of Variation :

وهو الشكل النسبي للانحراف المعياري ويعرف بالعلاقة :

$$CV = \frac{\sigma}{\bar{x}}.100\% \qquad (27\text{-}2)$$

وهو يعبر عن التشتت النسبي للظاهرة المدروسة وإن قيمته تكون كبيرة إذا كانت أكثر من50%، وهو يستخدم لمقارنة تباينات المتحول X بين المجموعات.

(2–8–2–4) : إنشاء مجالات الثقة :

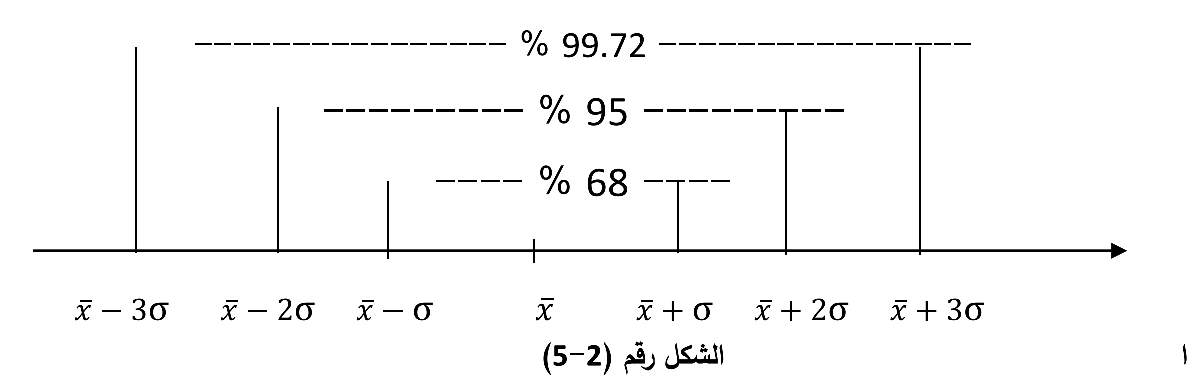
يستفاد من الانحراف المعياري في إنشاء مجالات الثقة المختلفة حول المتوسط $\bar{x}$ التالية :

مجال الثقة الأول : وهو المجال الذي يعتمد على العبارة $\bar{x} \pm \sigma$, أي أنه هو المجال: $[\bar{x} - \sigma \;,\; \bar{x} + \sigma]$ وهو يتضمن حوالي 68% فقط من القيم المتوفرة أو غير المتوفرة x_i .

مجال الثقة الثاني : وهو المجال الذي يعتمد على العبارة $\bar{x} \pm 2\sigma$, أي أنه هو المجال : $[\bar{x} - 2\sigma \;,\; \bar{x} + 2\sigma]$ وهو يتضمن حوالي 95% من القيم المتوفرة وغير المتوفرة x_i .

مجال الثقة الثالث : وهو المجال الذي يعتمد على العبارة $\bar{x} \pm 3\sigma$ أي أنه هو المجال : $[\bar{x} - 3\sigma \;,\; \bar{x} + 3\sigma]$ وهو يتضمن حوالي 99.72% من القيم المتوفرة وغير المتوفرة x_i .

ويمكننا تجسيد هذه المجالات على المحور OX كما يلي :



الشكل رقم (2−5)

ويستفاد من هذه المجالات في جميع مسائل التقدير واتخاذ القرارات.

مثال (2-16) : لنفترض إننا قمنا بقياس مستوى الكوليسترول العام Cholesterol لـ 500 طالباً، وقمنا بتبويبها في جدول كالتالي (mg/dl):

الجدول(2−15): بيانات الطلاب المبوبة مع بعض الحسابات اللازمة.

I	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	مجموع
مجالات قيم الكوليسترول x	120 – 130	130 – 140	140 – 150	150 – 160	160 – 170	170 – 180	180 – 190	190 – 200	200 – 210	210 – 220	220 – 230	230 – 240	240 – 250	250 – 260	260 – 270	
مراكز المجالات x_i'	125	135	145	155	165	175	185	195	205	215	225	235	245	255	265	
التكرارات المطلقة n_i	4	16	15	31	51	58	65	69	50	57	33	21	16	10	4	500
$n_i\ x_i'$	500	2160	2175	4805	8415	10150	12025	13455	10250	12255	7425	4935	3920	2550	1060	96080
$n_i\ x_i'^2$	62500	291600	315375	744775	1388475	1776250	2224625	2623725	2101250	2634825	1670625	1159725	960400	650250	280900	18885300

الحل: بعد إجراء الحسابات اللازمة لحساب المتوسط $\bar{x}$ والتباين σ^2 ، والتي وضعناها في السطرين الأخيرين للجدول السابق نجد أن المتوسط الحسابي لها يساوي:

$$\bar{x} = \frac{\sum ni\ x_i'}{n} = \frac{96080}{500} = 192.16$$

أما التباين فنحسبه من العلاقة المعدلة من العلاقة (4−7) السابقة وهي:

$$\sigma^2 = \frac{\sum n_i . x_i^2}{n} - (\bar{x})^2$$

ومن السطر الأخير في الجدول السابق نجد أن:

$$\sigma^{2} = \frac{18885300}{500} - (192.16)^{2} = 845.13$$

ومنها نجد أن الانحراف المعياري يساوي:

$$\sigma = \sqrt{845.13} = 29.07$$

وهكذا نجد أن معامل الاختلاف يساوي\:

$$CV = \frac{\sigma}{\bar{x}} = \frac{29.07}{192.16} 100 = 15\%$$

ولإنشاء مجالات الثقة نجد أن:

- المجال الأول للثقة وهو:

$$[\bar{x} - \sigma , \bar{x} + \sigma] = [192.16 - 29.07 , 192.16 + 29.07] = [163.09 , 221.23]$$

- المجال الثاني للثقة وهو:

$$[\bar{x} - 2\sigma , \bar{x} + 2\sigma] = [192.16 - 2(29.07) , 192.16 + 2(29.07)] = [134.02 , 250.31]$$

- المجال الثالث للثقة وهو:

$$[\bar{x} - 3\sigma , \bar{x} + 3\sigma] = [192.16 - 3(29.07) , 192.16 + 3(29.07)] = [104.95 , 279.38]$$

9-2 :العزوم المركزية: وسنقتصر هنا على تعريف العزمين الثالث والرابع وهما :

- العزم المركزي الثالث: وهو تعميم لفكرة التباين، وهو عبارة عن متوسط مكعبات انحرافات القيم عن متوسطها الحسابي $\bar{x}$، ونرمز له بـ M_3 ويعرف بالعلاقة التالية:

$$M_3 = \frac{\sum (x_i - \bar{x})^{3}}{n} \qquad (28-2)$$

ويتميز هذا العزم بأن قيمته يمكن أن تكون موجبة أو سالبة أو معدومة لأن اسه مفرد (يساوي 3).

فإذا كانت قيمته موجبة فإن هذا يحصل بسبب أن أكثر القيم تكون واقعة في جهة اليمين عن المتوسط الحسابي $\bar{x}$، وهذا يعني أن التوزيع التكراري لـ X يكون مائلاً إلى اليمين.

أما إذا كانت قيمته سالبة فإن هذا يحصل بسبب أن أكثر القيم تكون واقعة في جهة اليسار عن المتوسط الحسابي $\bar{x}$، وهذا يعني أن التوزيع التكراري لـ X يكون مائلاً إلى اليسار.

أما إذا كانت قيمته مساوية للصفر فإن هذا يحدث عندما تكون القيم موزعة بالتساوي تقريباً على جانبي المتوسط الحسابي $\bar{x}$، وهذا يعني أن التوزيع التكراري لـX يكون متناظراً أو شبه متناظر حول المتوسط $\bar{x}$ ولهذا فإنه يستفاد من هذا العزم في تعريف التناظر وقياس مقدار الالتواء كما سنرى لاحقاً.

- العزم المركزي الرابع: وهو أيضاً تعميم لفكرة التباين، وهو عبارة عن متوسط الأس الرابع لانحرافات القيم عن متوسطها الحسابي $\bar{x}$ ، ونرمز له بـM4 ويعرف بالعلاقة:

$$M4 = = \frac{\sum (x_i - \bar{x})^{4}}{n} \qquad (29-2)$$

وإن هذا العزم يأخذ قيمة موجبة دوماً (لأن أسه 4 زوجي) ويستخدم في قياس تطاول التوزيعات التكرارية للمتحولات العشوائية.

2-10 : تطبيقات الانحراف المعياري والعزوم المركزية:

يستفاد من الانحراف المعياري والعزوم المركزية في تعريف مقاييس جديدة لدراسة مقدار التواء (ميلان) أو تطاول التوزيع التكراري كما يلي:

- مقياس الالتواء (الميلان) Skewness :

لقد أشرنا سابقاً إلى أنه يمكننا الاستفادة من العزم المركزي الثالث M_3 في تعريف وقياس مقدار التواء التوزيعات التكرارية للمتحولات العشوائية، وذلك بتعريف مؤشر نسبي له وحسابه من العلاقة التالي:

$$K = \frac{M_3}{\sigma^3} = \frac{\left(\frac{1}{n}\sum(x_i - \bar{x})^3\right)}{\sigma^3} \qquad (30-2)$$

- فإذا كانت قيمة هذا المقياس $K > 0$ فإن هذا يعني أن التوزيع التكراري يكون ملتوياً نحو اليمين ويكون الالتواء كبيراً نحو اليمين كلما كانت قيمة K كبيرة.
- أما إذا كانت قيمة هذا المقياس $K < 0$ فإن هذا يعني أن التوزيع التكراري يكون ملتوياً نحو اليسار ويكون الالتواء كبيراً نحو اليسار كلما كانت قيمة K السالبة كبيرة.
- أما إذا كانت قيمة $K = 0$ أو كانت قريبة من الصفر فإن هذا يعني أن التوزيع التكراري يكون متناظراً حول المتوسط $\bar{x}$ أو شبه متناظر حوله.

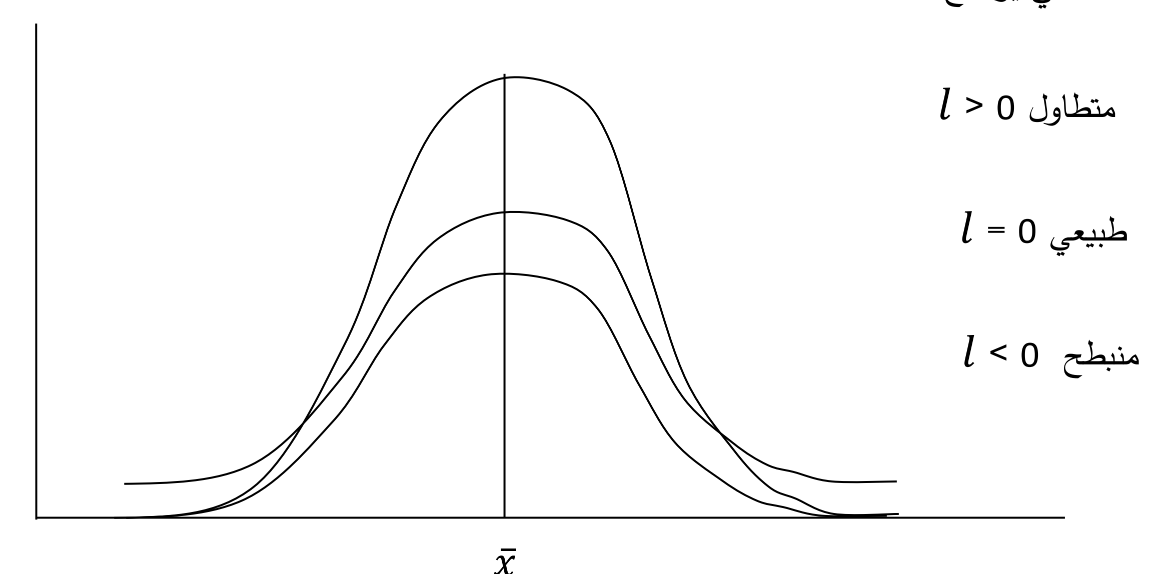
- مقياس التطاول Kurtosis (tailedness) :

وهو يعتمد على العزم المركزي الرابع M_4 وعلى مربع التباين ويعرف بالعلاقة التالية:

$$l = \frac{M_4}{\sigma^4} - 3 \qquad (31-2)$$

- فإذا كانت قيمته $l > 0$ فإن هذا يعني أن تطاول التوزيع التكراري أكبر من التطاول الطبيعي ، وتزداد قيمته كلما كان التطاول كبيراً.
- أما إذا كانت قيمته $l < 0$ فإن هذا يعني أن تطاول التوزيع التكراري أصغر من التطاول الطبيعي ، وعندها يكون التوزيع منبطحاً .وتتناقص قيمة l كلما كان التطاول صغيراً.
- أما إذا كانت قيمته $l = 0$ أو قريبة من الصفر ، فإن هذا يعني أن تطاول التوزيع التكراري يكون طبيعياً أو شبه طبيعي.

والشكل التالي يوضح هذه الحالات :



الشكل رقم (2− 6): أشكال التطاول

مثال (2-17) : لنفترض أننا أخذنا النتائج المخبرية لقياس مستويات الهيموغلوبين في الدم لعينة من الطلاب الأصحاء بحجم n = 100 ، والتي نفترض أنها كانت مبوبة كما يلي (g/dl):

الجدول(2−16): بيانات الطلاب المبوبة.

I	1	2	3	4	5	6	المجموع $\sum$
مجالات التبويب $[x_i , x_{i+1}[$	[13,15[	[15, 17[	[17,19[	[19,21[	[21,23[	[23,29[	
التكرارات n_i المطلقة	0	10	30	40	10	10	100

والمطلوب حساب العزوم المركزية ومقاييس الالتواء والتطاول للتوزيع التكراري لمستوى الهيموغلوبين X .

الحل: نقوم أولاً بحساب المتوسط الحسابي $\bar{x}'$ وانحرافات القيم عنه, ثم نربعها ونحسب التباين σ^2 والانحراف المعياري σ ، ثم نأخذ مكعباتها ثم نحسب أسها الرابع ثم نتابع الحسابات اللازمة. ولإجراء هذه الحسابات المعقدة نعد جدولاً خاصاً لذلك. فنحصل على ما يلي:

الجدول(2−17):الحسابات المساعدة

مراكز المجالات x'_i	14	16	18	20	22	24	المجموع
الجداءات $n_i\, x'_i$	0	160	540	800	220	240	1960
الانحرافات $(x'_i - \bar{x})$	−3.6	−3.6	−1.6	0.4	2.4	4.4	X
$n_i(x'_i - \bar{x})^2$	0	129.6	76.8	6.4	57.6	193.6	464.00
$n_i(x'_i - \bar{x})^3$	0	−466.56	−122.88	2.56	138.24	851.784	+403.20
$n_i(x'_i - \bar{x})^4$	0	6.1679	196.6	1.024	331.8	3748.09	5957.12

ومن هذا االجدول نجد أن المتوسط الحسابي يساوي :

$$\bar{\mathbf{x}} = \frac{\sum n_i x'_i}{\mathbf{n}} = \frac{\mathbf{1960}}{\mathbf{100}} = \mathbf{19.6}\ \boldsymbol{g/dl}$$

وأن التباين والانحراف المعياري يساويان :

$$\boldsymbol{\sigma^2} = \frac{\sum \mathbf{n_i}\left(x'_i - \bar{\mathbf{x}}\right)^2}{\mathbf{n}} = \frac{\mathbf{464}}{\mathbf{100}} = 4.64$$

$$\sigma = \sqrt{464} = 2.154$$

وإن العزم المركزي الثالث يساوي :

$$M_3 = \frac{\sum \mathbf{n_i}\left(x'_i - \bar{\mathbf{x}}\right)^3}{\mathbf{n}} = \frac{\mathbf{403.20}}{\mathbf{100}} = 4.032$$

وإن العزم المركزي الرابع يساوي :

$$M_4 = \frac{\sum \mathbf{n_i}\left(x'_i - \bar{\mathbf{x}}\right)^4}{\mathbf{n}} = \frac{\mathbf{5957.12}}{\mathbf{100}} = 59.57$$

وإن مقياس الالتواء K يساوي :

$$K = \frac{\boldsymbol{M_3}}{\sigma^3} = \frac{\mathbf{4.032}}{(\mathbf{2.154})^3} = \frac{\mathbf{4.032}}{\mathbf{9.99}} = 0.40$$

وهو مقدار موجب وصغير، وهذا يعني أن التوزيع ملتو قليلاً إلى اليمين, ولكنه قريب إلى الشكل الطبيعي المتناظر.

وإن مقياس التطاول l يساوي:

$$l = \frac{\boldsymbol{M_4}}{\sigma^4} - 3 = \frac{\mathbf{59.57}}{(\mathbf{2.154})^4} - 3 = \frac{\mathbf{59.57}}{\mathbf{21.53}} - 3 = -0.233$$

وهو مقدار سالب وصغير، وهذا يعني أن تطاول التوزيع التكراري لهذه العينة أقل من التطاول الطبيعي بقليل، أي أن التوزيع شبه طبيعي.

تمـريـنات

1– لتكن لدينا القياسات التالية عن مستوى اللمفاويات في الدم (mg/dl):

X :	20	21	22	24	26	28	30	33	34	35	36
	36	37	38	39	40	24	24	34	35	32	31

والمطلوب :

- ترتيب هذه القياسات تصاعدياً وحساب التكرارات المقابلة لقيمتها.
- حساب المتوسط الحسابي والوسيط والمنوال لهذه القياسات.
- حساب المدى والتباين والانحراف المعياري وحساب مجالات الثقة لها.
- حساب العزمين الثالث والرابع.
- حساب مقياس الالتواء والتطاول.
- تجزئة هذه القياسات إلى مجموعتين: الأولى من]30– 20] والثانية من]40– 30] ، ثم حساب المتوسط الحسابي لكل منها، ثم حساب التباين الداخلي ضمن كل منها، ثم حساب التباين بين المجموعتين، ثم التأكد من أن التباين الكلي السابق يساوي مجموع متوسط التباينات الداخلية والتباين بين المجموعتين.

2– لنفترض أنه أخذنا قياسات كمية الكوليسترول في الدم (mg/dl) لعينة من الطلاب الأصحاء بحجم n=80 ، وبوبناها فكانت كما كما يلي:

I	1	2	3	4	5	6	المجموع
مجالات التبويب $[x_i , x_{i+1}[$	120 , 150	150 , 170	170 , 190	190 , 210	210 , 230	230 , 250	×
التكرارات n_i المطلقة	8	16	25	17	8	6	80

والمطلوب حساب مايلي:

- المتوسط الحسابي والتباين والانحراف المعياري ومعامل الاختلاف لكمية الكوليسترول في الدم.
- مجالات الثقة للمتوسط: الأول والثاني والثالث.
- الوسيط والمنوال.
- العزمين الثالث والرابع.
- مقياسي الالتواء والتطاول.

الفصل الثالث

المتحولات العشوائية والتوزيعات الاحتمالية

3-0: مفهوم المتحولات العشوائية:

يمكننا ببساطة أن نعرف المتحولات العشوائية بأنها هي المتحولات التي تأخذ قيمها الممكنة بصورة عشوائية وغير معروفة مسبقاً وبتكرارات معينة. وهي قد تكون مستمرة أو منقطعة، ولتوضيح ذلك نقدم الأمثلة التالية:

وزن الطفل عند الولادة: وهو متحول عشوائي مستمر ويقاس بالغرام، ويتحول ضمن مجال معين(مثل من2000 إلى 5000 غرام)، ويتغير من طفل لآخر بصورة عشوائية وغير معروفة مسبقاً, ولكنه يأخذ قيمه الممكنة المحددة في المجال السابق بتكرارات معينة. ويمكن حساب هذه التكرارات وحساب التكرارات النسبية واستخلاص التوزيع التكراري منها.

عدد أفراد الأسرة: وهو متحول عشوائي منقطع ويقاس بالفرد، ويتحول ضمن نطاق عددي معيّن **(**من 1، 2، 3،.... 15) ويأخذ قيمه الممكنة بتكرارات معينة. ويمكننا حساب هذه التكرارات واستخلاص التوزيع التكراري منها.

عمر الطفل عند ظهور أسنانه اللبنية: وهو متحول عشوائي مستمر ويقاس بالأسابيع أو الأشهر، وهو يختلف من طفلٍ لآخر ويأخذ قيمه الممكنة ضمن مجالٍ معين بصورة عشوائية وغير معروفة مسبقاً، وبتكرارات معينة. ويمكننا حساب هذه التكرارات واستخلاص التوزيع التكراري منها.

عمر الطالب في الجامعة: وهو متحول عشوائي مستمر يختلف من طالبٍ لآخر ويأخذ قيمه الممكنة بصورة عشوائية وغير معروفة مسبقاً وبتكرارات معينة، ويمكننا حساب هذه التكرارات واستخلاص التوزيع التكراري منها.

نستنتج من الأمثلة السابقة أن المتحولات العشوائية تنقسم إلى نوعين أساسين: متحولات مستمرة أو متحولات منقطعة. كما نلاحظ أنه لابد أن يكون لكل متحول عشوائي عنصران أساسيان هما:

. مجموعة من القيم الممكنة المنقطعة أو المستمرة ضمن مجال معين (فضاء).

. توزيع احتمالي يعبّر عن الاحتمالات المقابلة لكل قيمة من القيم الممكنة.

ومن خلال معرفتنا لهذين العنصرين يمكن استخلاص العديد من النتائج وحساب جميع المميزات العددية للمتحول العشوائي: كالمتوسط والتباين والانحراف المعياري ... إلخ.

وهناك نوعان من التوزيعات الاحتمالية هما:

- **التوزيعات الاحتمالية النظرية:** وهي التوزيعات التي تعطى بواسطة معادلة رياضية معرفة على مجال القيم الممكنة. وتعطينا الاحتمالات المقابلة لكل قيمة أو مجال من قيمه الممكنة.

- **التوزيعات الاحتمالية التجريبية:** وهي عبارة عن التوزيعات التكرارية (التكرارات النسبية) المقابلة لجميع القيم الممكنة للمتحول العشوائي المدروس, والتي نحصل عليها من خلال الدراسات أو التجارب الإحصائية.

وتتميز كل هذه التوزيعات بأن الاحتمالات فيها غير سالبة وإن مجموعها يساوي الواحد.

وتصنف التوزيعات الاحتمالية النظرية والتجريبية إلى نوعين أساسين هما:

- **التوزيعات الاحتمالية المنقطعة:** وهي التوزيعات الخاصة بالمتحولات المنقطعة مثل عدد أفراد الأسرة، عدد الطلاب، عدد المرضى... إلخ.
- **التوزيعات الاحتمالية المستمرة:** وهي التوزيعات الخاصة بالمتحولات المستمرة مثل: درجة الحرارة، عمر الطالب، وزن الطفل، طول الشخص ... إلخ.

3-1: بعض التوزيعات الاحتمالية المنقطعة:

يكون المتحول العشوائي **X** منقطعاً عندما يأخذ قيماً منقطعة x_i (صغيرة), ونكتب ذلك كما يلي:

X : $x_1 \quad x_2 \quad x_3 \quad .... x_i .... x_n$

ولتوضيح فكرة التوزيعات الاحتمالية المنقطعة للمتحولات العشوائية المنقطعة نأخذ الأمثلة التالية:

- نتائج حجر النرد تأخذ القيم التالية : **X:** 1, 2, 3, 4, 5, 6
- عدد أفراد الأسرة يأخذ القيم التالية: **X:** 1,2, 3, 4, 5, 6, 7, 8, 9, 10, 11..
- عدد زوار الطبيب خلال اليوم : يمكن أن يأخذ القيم التالية: **X:** 1,2, 3, 4, 5, 6, 7, 8
- عدد المرضى الذين يستقبلهم قسم الإسعاف خلال ساعة ،يمكن أن يأخذ القيم التالية:

X : 0, 1, 2, 3, 4, 5, 6, 7, 8**......**

وبما أن هذه المتحولات هي متحولات عشوائية فإنها تأخذ كل من قيمها الممكنة بتكرارات معينة.

لذلك يكون لكل متحول عشوائي منقطع توزيع احتمالي منقطع، بحيث يقابل كل قيمة x_i من قيم **X** تكرار معين أو احتمال محدد P_i ، وتكون هذه الاحتمالات غير سالبة ويكون مجموعها مساوياً للواحد ($\sum P_i = 1$), ويمكننا أن نكتب ذلك التوزيع الاحتمالي ضمن جدول مناسب كما يلي:

قيم X	x_1	x_2	x_3	x_4	x_i	x_n	$\sum$
الاحتمالات المقابلة P_i	P_1	P_2	P_3	P_4	P_i	P_n	1

كما يمكننا أن نحسب متوسط قيم **X** (التوقع الرياضي لها) من العلاقة :

$$E(X) = \sum p_i x_i \qquad (1\text{-}3)$$

وأن نحسب تباينها وانحرافها المعياري من العلاقتين :

$$\sigma^2_{(X)} = \sum p_i (x_i - E)^2 = \sum p_i x_i^2 - E^2 \quad \rightarrow \quad \sigma = \sqrt{\sigma^2} \qquad (2\text{-}3)$$

وبما أنه يوجد لدينا العديد من المتحولات العشوائية المنقطعة التي تخضع لتوزيعات احتمالية متشابهة, فإنه يمكن التعبير عنها بتوزيع نظري موحد, ولذلك سنستعرض بعض التوزيعات النظرية المنقطعة التالية:

3‒1‒1: التوزيع المنتظم: uniform distribution

وهو التوزيع الاحتمالي الذي تكون فيه الاحتمالات المقابلة لقيم **X** متساوية مثل :

توزيع نتيجة حجر النرد: وهو يأخذ الشكل التالي:

قيم X	1	2	3	4	5	6	$\sum$
التوزيع الاحتمالي P_i	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

توزيع نتيجة سحب الأرقام في عجلة الرهان (اليانصيب): وهو يأخذ الشكل التالي:

قيم X	0	1	2	3	4	5	6	7	8	9	$\sum$
التوزيع الاحتمالي P_i	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	1

توزيع نتيجة الولادة: وهو يأخذ الشكل التالي:

قيم X	ذكر	انثى	$\sum$
التوزيع الاحتمالي P_i	$\frac{1}{2}$	$\frac{1}{2}$	1

وبصورة عامة يمكننا أن نعرف التوزيع النتقطع المنتظم لجميع هذه المتحولات العشوائية المنتظمة بعلاقة موحدة هي العلاقة التالية:

$$\boldsymbol{P_i = \frac{1}{n} \qquad 1 \le i \le n \qquad (3-3)}$$

حيث **i** هو الدليل العددي لقيم **X** , و **n** العدد الكلي لقيم **X** .

وبناءً عليه يمكننا حساب التوقع الرياضي لهذه المتحولات المنتظمة وتباينها وانحرافها المعياري كما يلي:

$$E(\mathbf{X}) = \sum \mathrm{P}_i \,.\, x_i = \sum \frac{1}{n} x_i = \frac{1}{n} \sum x_i = \bar{x} \qquad (4-3)$$

$$\sigma^2_{(\mathbf{X})} = \sum P_i (x_i - E)^2 = \sum \frac{1}{n} (x_i - \bar{x})^2 = \sigma^2 \qquad (5-3)$$

$$\sigma = \sqrt{\sigma^2} \qquad (6-3)$$

فمثلا نجد أن التوقع الرياضي لنتيجة حجر النرد يساوي:

$$E(\mathbf{X}) = \frac{1}{6}.1 + \frac{1}{6}.2 + \frac{1}{6}.3 + \frac{1}{6}.4 + \frac{1}{6}.5 + \frac{1}{6}.6 = 3.5$$

وكذلك نجد أن التباين والانحراف المعياري يساويان:

$$\sigma^2_{(x)} = \frac{1}{6}(1-3.5)^2 + \frac{1}{6}(2-3.5)^2 + \frac{1}{6}(3-3.5)^2 + \frac{1}{6}(4-3.5)^2 + \frac{1}{6}(5-3.5)^2 + \frac{1}{6}(6-3.5)^2$$

$$\sigma^2_{(x)} = 2.917$$

$$\sigma = \sqrt{2.917} = 1.708$$

ملاحظة : يلعب هذا التوزيع دوراً كبير في سحب الأرقام العشوائية, ويعتمد عليه في إيجاد حلول مثالية لكثير من القضايا الاجتماعية والطبيعية.

2–1–3 : التوزيع الثنائي (ذو الحدين): Binomial Distribution :

يستخدم هذا القانون للتعبير عن الاحتمالات المتعلقة بالظواهر الثنائية (التي لها وجهان فقط: نجاح وفشل), ويعتمد عليه في حساب الاحتمالات المقابلة لظهور الوجه المطلوب عدداً من المرات يساوي ***k*** مرة تماماً عند تكرار التجربة عليها ***n*** مرة .

ويشترط هنا أن يكون احتمال تحقق الوجه المطلوب في كل التجارب ثابتاً ومعلوماً ويساوي ***P*** .

وكمثال على ذلك سنقوم بحساب الاحتمال المقابل لولادة /3/ ذكور تماماً من أصل/4/ ولادات . علماً بأن احتمال ولادة الذكر في كل ولادة ثابتٌ ونعتبره مساوياً $\left(\frac{1}{2}\right)$.

أو بشكل عام نرمز لاحتمال ولادة الذكر بالرمز ***P***، ولاحتمال ولادة الأنثى بالرمز **q=1-p** .

ولنرمز لعدد الذكور, الذي يمكن أن ينتج عن هذه الولادات الأربعة بالرمز ***X*** , وهنا نلاحظ أن ذلك المتحول العشوائي ***X*** يمكن أن يأخذ القيم التالية : **X : 0, 1, 2, 3, 4** *, وسنعمل على إيجاد التوزيع الاحتمالي لهذا المتحول. لذلك نرمز لحادث ولادة الذكر بالرمز* **M** ولحادث ولادة الأنثى بـ **F** , فنجد أن هذه الولادات الأربعة يمكن أن تعطينا أحدى النتائج الممكنة التالية (المرتبة حسب تسلسل النوع):

عدد الذكور أو قيم X	النتائج الممكنة	الاحتمال المقابل
عدم ولادة أي ذكر 0	F.F.F.F	$q.q.q.q.=q^4$
ولادة ذكر واحد 1	MFFF,FMFF,FFMF, FFFM	$4.P.q.q.q = 4pq^3$
ولادة ذكرين 2	MMFF,MFMF,FMFM ,FFMM,FMMF MFFM	$6.p.p.q.q = 6p^2q^2$
ولادة ثلاثة ذكور 3	MMMF,MMFM,MFMM,FMMM	$4.P.P.P.q = 4p^3.q$
ولادة أربعة ذكور 4	MMMM	$p.p.p.p = p^4$

وهكذا نجد أن التوزيع الاحتمالي لعدد الذكور***X*** نتيجة هذه الولادات الأربعة يأخذ الشكل التالي :

قيم *X*	0	1	2	3	4	$\sum$
التوزيع الاحتمالي	q^4	$4Pq^3$	$6P^2q^2$	$4P^3 q$	P^4	1

وهنا نلاحظ أن الأعداد التي تسبق الرموز السابقة هي عبارة عن عدد المتوافقات الممكنة للنتيجة المقابلة لها, والتي يرمز بـ C_n^k.

وبصورة عامة يمكننا التعبير عن احتمال تحقق الوجه المرغوب (k مرة تماماً) عند إجراء التجربة على هذه الظاهرة الثنائية n مرة, وباحتمال ثابت P في كل تجربة, بواسطة العلاقة التالية :

$$P(X = k) = C_n^k \,.\, P^k \,.\, q^{n-k} \qquad (7-3)$$

حيث أن: k=0,1,2,3,......n , وأن: C_n^k هو عدد المتوافقات بحجم k من n عنصراً ويساوي:

$$C_n^k = \frac{n!}{k!(n-k)!} \qquad (8-3)$$

ويمكن البرهان على أن التوقع الرياضي للعدد X الخاضع للتوزيع الثنائي يساوي:

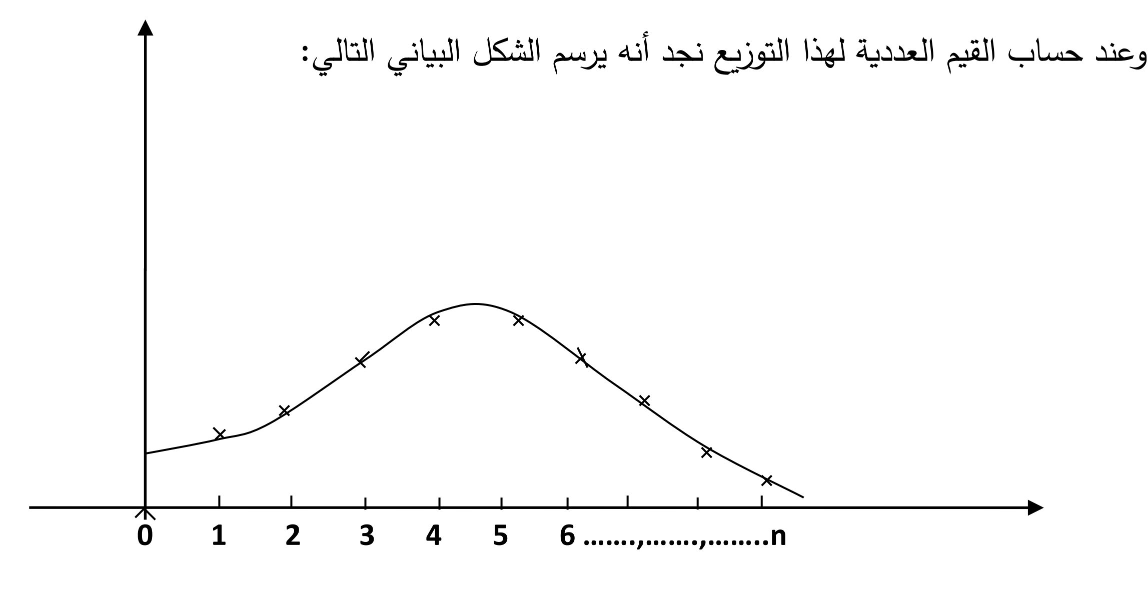
$$E(X) = n \,.\, P \qquad (9-3)$$

وعلى أن التباين والانحراف المعياري له يساويان:

$$\sigma^2_{(X)} = n \,.\, P \,.\, q \qquad (10-3)$$

$$\sigma_{(X)} = \sqrt{n \,.\, P \,.\, q} \qquad (11-3)$$

وعند حساب القيم العددية لهذا التوزيع نجد أنه يرسم الشكل البياني التالي:



الشكل (3–1): الشكل العام للتوزيع الثنائي (عندما $P = \frac{1}{2}$)

ملاحظة: يمكن تطبيق هذا القانون على معظم الظواهر بعد تحويلها إلى ظواهر ثنائية ، كأن نأخذ أحد أو بعض الحالات ونعتبرها الوجه المطلوب فتكون الحالات المتممة هي الوجه الثاني. وذلك بشرط أن يكون احتمال تحقق الوجه المطلوب ثابتاً ويساوي p . وكمثال على ذلك يمكن أن نجعل نتيجة حجر النرد ظاهرة ثنائية بأن نعتبر مجموعة الأوجه الفردية هي الوجه المطلوب ويكون احتمال تحققه في كل تجربة ثابتاً ويساوي $p=3/6$.

مثال (3−1) : إذا كانت نسبة العطب في إنتاج إحدى شركات الأدوية **P= 0.09**. وإذا اشترينا بشكل عشوائي /5/ قطع من منتجاتها. فأوجد التوزيع الاحتمالي لعدد القطع, التي يمكن أن تكون معطوبة في هذه الصفقة.

الحل : لنرمز لعدد القطع, التي يمكن أن تكون معطوبة في هذه الصفقة بـ ***X*** , فنجد أن ***X*** يمكن أن يأخذ القيم التالية:

X= 0, 1, 2, 3, 4, 5

وبما أن هذه الظاهرة ثنائية وإن الوجه المطلوب منها هو وجه العطب وإن احتماله في المصنع ثابت ويساوي ***P=0.09***, فإن التوزيع الاحتمالي لهذه الحالة هو التوزيع الثنائي المعرف بالعلاقة:

$$P(X=k)=C_5^k\,P^k\,q^{5-k}$$

حيث أن: K = 0 1 2 3 4 5

وبناء على ذلك نحسب التوزيع الاحتمالي المفصل كما يلي:

$$P(X=0)=C_5^0\,.P^0\,.q^5=\ q^5=(1-0.09)^5=0.6240$$

$$\mathrm{P(X=1)=C_5^1\,.P^1\,.q^4}=5(0.09)(1-0.09)^4=0.3086$$

$$\mathrm{P(X=2)=C_5^2\,.P^2\,.q^3}=10(0.09)^2(1-0.09)^3=0.0610$$

$$\mathrm{P(X=3)=C_5^3\,.P^3\,.q^2}=10(0.09)^3(1-0.09)^2=0.0060$$

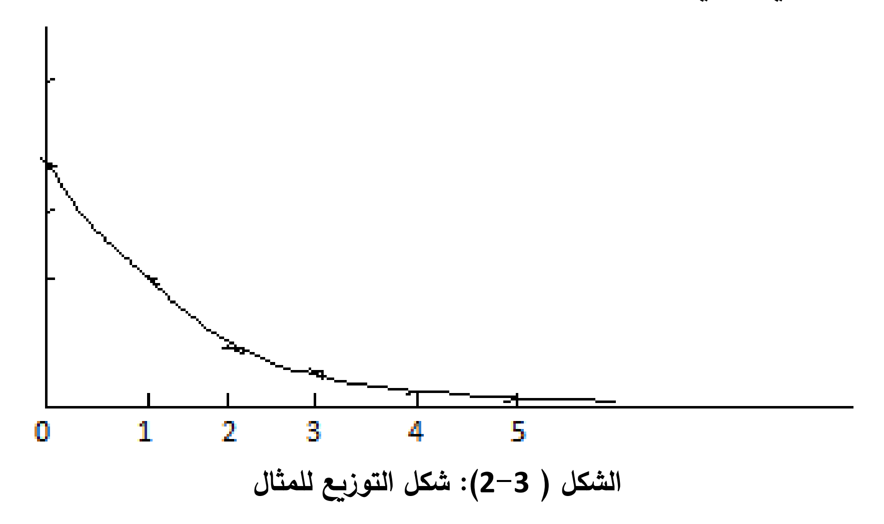
$$\mathrm{P(X=4)=C_5^4\,.P^4\,.q^1}=5(0.09)^4(1-0.09)^1=0.0003$$

$$\mathrm{P(X=5)=C_5^5\,.P^5\,.q^0}=P^5=(0.09)^5=0.0000$$

ثم نضع النتائج في جدول كالتالي:

عدد القطع المعطوبة أو قيم *X*	0	1	2	3	4	5	∑
التوزيع الاحتمالي	0.6240	0.3086	0.0610	0.0060	0.0003	0.0000	1

وهو يرسم الشكل البياني التالي :



الشكل (3-2): شكل التوزيع للمثال

3-1-3: توزيع بواسون: Poisson Distribution :

وهو يعالج الاحتمالات المتعلقة بالظواهر الثنائية النادرة (التي لها وجهان : نجاح وفشل)، ويستخدم لحساب الاحتمالات المقابلة لظهور الوجه النادر عدداً k تماماً من المرات عند إجراء وتكرار التجربة عدداً لا نهائياً من المرات . وهو يعطى بالعلاقة التالية :

$$p(x = k) = \frac{\lambda^k}{k!}\bar{e}^{\lambda} \qquad (\lambda > 0) \quad (12-3)$$

حيث **k** تأخذ القيم التالية : **k : 0, 1, 2, 3, ...** ∞

ويشترط عند تطبيق هذا القانون أن يكون احتمال الوجه النادر هو الوجه المطلوب وأن يكون احتماله ***p*** ثابتا وصغيراً جداً ، وأن يكون الجداء **(n.p)** ثابتاً ويساوي **λ** , أي أن يكون:

$$\lambda = n\,.\,p \qquad (13-3)$$

ويمكن البرهان على أن التوقع الرياضي للمتحول ***x*** الخاضع لهذا التوزيع يساوي:

$$E\,(x) = \lambda \qquad (14-3)$$

وإن تباينه وانحرافه المعياري يساويان :

$$\sigma^2_{(x)} = \lambda \qquad (15-3)$$

$$\sigma = \sqrt{\lambda}$$

مثال (3-2): أوجد التوزيع الاحتمالي لعدد القطع المعطوبة عند شراء /5/ قطع من منتجات شركة الأدوية المذكورة في المثال السابق.

الحل : لدينا **p = 0.09** ، وهو احتمال العطب وله قيمة ثابتة وصغيرة نسبياً، لذلك يمكننا حساب قيم التوزيع الاحتمالي اعتماداً على توزيع بواسون المعرف بالعلاقة (3-12)، بعد أن نحسب قيمة λ من العلاقة :

$$\lambda = n.p = 5.(0.09) = 0.45$$

وبذلك نجد أن :

$$p(x=0)=\frac{(0.45)^0}{0!}\bar{e}^{0.45}=0.637628$$

$$p(x=1)=\frac{(0.45)^1}{1!}\bar{e}^{0.45}=0.286933$$

$$p(x=2)=\frac{(0.45)^2}{2!}\bar{e}^{0.45}=0.0645598$$

$$p(x=3)=\frac{(0.45)^3}{3!}\bar{e}^{0.45}=0.00968398$$

$$p(x=4)=\frac{(0.45)^4}{4!}\bar{e}^{0.45}=0.00108945$$

$$p(x=5)=\frac{(0.45)^5}{5!}\bar{e}^{0.45}=0.000098$$

وهنا نلاحظ أن هذه الاحتمالات قريبة جداً من الاحتمالات التي حسبناها من التوزيع الثنائي.

نتيجة هامة: يمكن استخدام توزيع بواسون لحساب قيم تقريبية لاحتمالات التوزيع الثنائي وذلك عندما يكون احتمال التحقق العام لظهور الوجه المطلوب ***p*** صغيراً جداً وعندما يكون عدد التجارب***n*** كبيراً جداً.

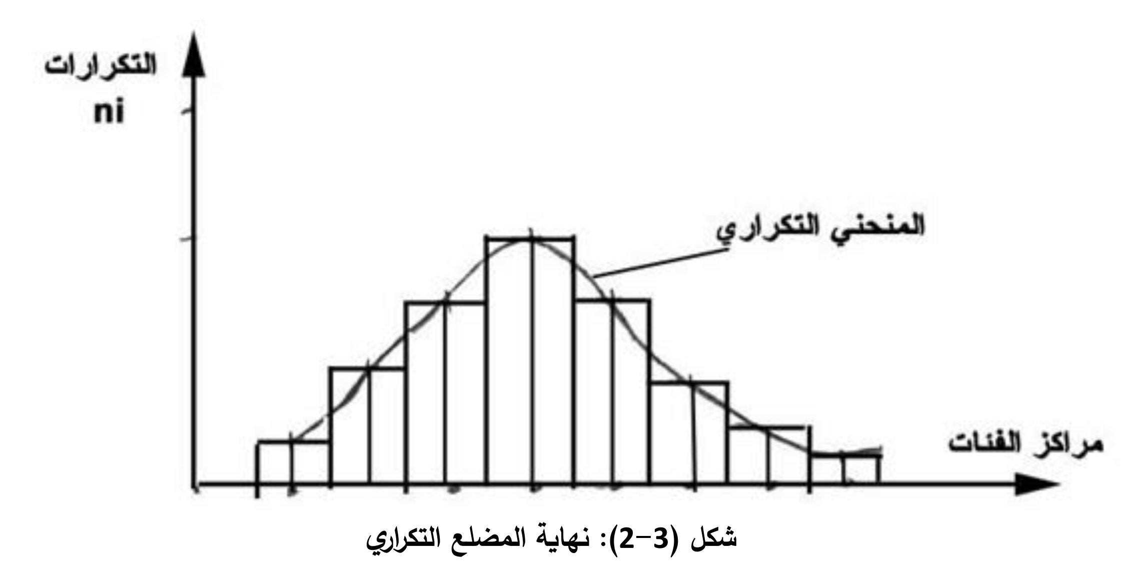
وأخيراً نشير إلى أنه يمكن حساب قيم الاحتمالات التقريبية للتوزيع الثنائي عندما تكون قيمة احتمال التحقق العام ***p*** غير صغيرة بواسطة التوزيع الطبيعي المعياري كما سنرى لاحقاً.

2-3 : بعض التوزيعات الاحتمالية المستمرة:

وهي التوزيعات الخاصة بالمتحولات المستمرة أو بالمتحولات التي يمكن اعتبارها مستمرة، مثل درجة الحرارة، دخل الأسرة، درجات الطالب، عمر الطالب...إلخ.

وإذا قمنا بتبويب المعلومات الإحصائية المتوفرة عن أي متحول مستمر **X** ضمن مجالات محددة, فإن التمثيل البياني للتكرارات النسبية المقابلة لها بواسطة أعمدة تأخذ شكلاً معيناً متناظراً أو مائلاً إلى اليمين أو مائلاً إلى اليسار.

وكمثال على ذلك نأخذ التوزيع التكراري لمستوى الشحوم الثلاثية لدى البالغين والذي يأخذ بعد تبويبه ضمن مجالات محددة شكلاً متناظراً كما في الشكل التالي:

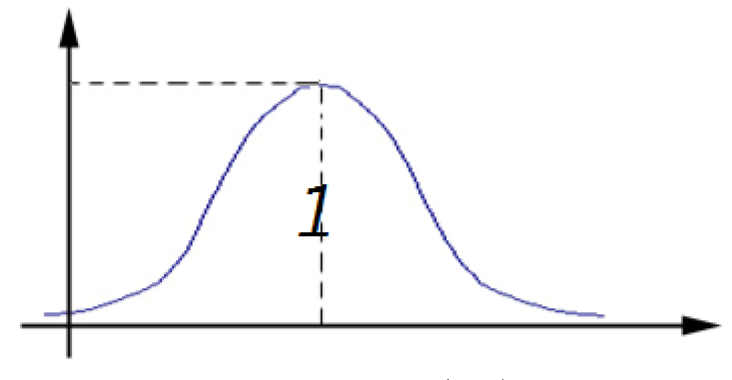


شكل (3–2): نهاية المضلع التكراري

وهنا نلاحظ أن احتمال وقوع المتحول العشوائي **X** في المجال الجزئي **]a, b[** يساوي مساحة العمود التكراري المرسوم فوق ذلك المجال. وأن مجموع مساحات هذه الأعمدة يساوي الواحد.

وإذا جعلنا المجالات الجزئية صغيرة جداً وجعلنا حجم العينة **n** كبيراً, فإن أعمدة التوزيع الاحتمالي تصبح على شكل أشرطة عمودية متجاورة.

وإذا قمنا بوصل قمم الأشرطة فإننا سنحصل على منحنٍ انسيابي يسمى منحني التوزيع الاحتمالي للمتحول العشوائي **X**، وهو سيأخذ الشكل الآتي:



شكل (3–3): منحني التوزيع الاحتمالي

وبما أننا نعتبر المتحول العشوائي **X** متحولاً مستمراً، فإنه يمكننا أن ننتقل إلى مرحلة جديدة ونعبّر عن هذه المنحنيات بواسطة معادلات رياضية معينة وذات متحولات مستمرة.

ولقد توصل الإحصائيون إلى صياغة عدد من المعادلات التي تصلح للتعبير عن الكثير من التوزيعات الاحتمالية. ولكن بما أن هذه المعادلات تتميّز ببعض التعقيد قد لا يستوعبه الدارسون في هذه المرحلة، فإننا سنكتفي بإيراد بعض الصيغ الرياضية والأشكال العامة لمنحنيات أهمّ التوزيعات الاحتمالية المستمرة.

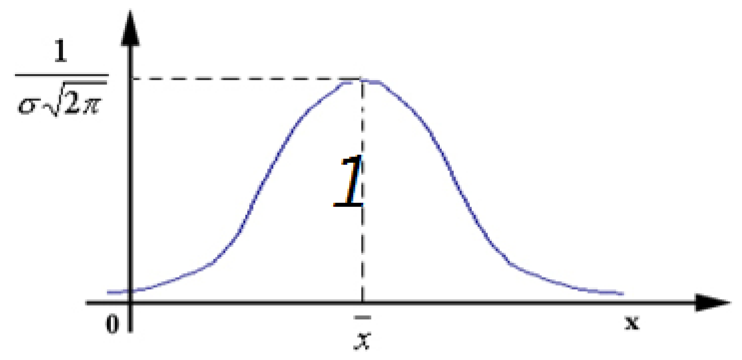
1–2–3 : التوزيع الطبيعي العام: Normal Distribution

ويعتبر هذا التوزيع أهم التوزيعات الاحتمالية وأكثرها تطبيقاً وانتشاراً، ويصلح لمعظم المتحولات العشوائية، وهو يعرف بالعلاقة التالية:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}.e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma}\right)^2} \qquad -\infty < X < +\infty \qquad (16-3)$$

حيث $\bar{x}$ هو المتوسط الحسابي أو التوقع، و σ الانحراف المعياري لـx .

وهو يرسم المنحنى التالي، والذي يسمى بمنحنى التوزيع الطبيعي العام أو منحنى (غاوص).



الشكل رقم (3–4): الشكل العام للتوزيع الطبيعي

ومن خصائص هذا التوزيع:

. إن توقعه هو المتوسط $\bar{X}$، وأن تباينه يساوي σ^2. ولذلك يرمز له بالرمز $N(\bar{x},\sigma^2)$.

. إن متناظر بالنسبة للمتوسط أو المركز $\bar{X}$.

. إن المساحة تحته تساوي الواحد.

. يقع بكامله فوق محور السينات (المحور الأفقي).

ومن المتحولات الكثيرة التي تخضع لهذا التوزيع نذكر المتحولات التالية:

درجة الطالب في الامتحان: حيث يفترض أنها تتوزع بشكل متناظر حول متوسط معين.

طول المولود أو وزنه عند الولادة: حيث يفترض أنه يتوزع بشكل متناظر حول متوسط معين.

نسبة الذكاء لطلاب مرحلة معينة: حيث يفترض أنها تتوزع بشكل متناظر حول متوسط معين.

مدة حياة الإنسان: حيث يفترض أنها تتوزع بشكل متناظر حول متوسط معين.

حجم المبيعات اليومي في إحدى الصيدليات... إلخ.

ولحساب احتمال أن يأخذ X قيمة أصغر أو تساوي x_1 يجب أن نقوم بإجراء التكامل التالي:

$$P(X \le X_1)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x_1} e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma}\right)^2}\cdot dx \qquad (17-3)$$

ومع أن هذا التوزيع هو من أكثر التوزيعات الاحتمالية انتشاراً، إلا أن عملية حساب الاحتمالات فيه اصطدمت بعدة عوائق رياضية وحسابية. ولقد تم التغلب على هذه العوائق بالاعتماد على حالة خاصة منه تسمى: (قانون التوزيع الطبيعي المعياري), لأنها تستخدم كمعيار عند حساب الاحتمالات.

3-2-2: التوزيع الطبيعي المعياري:Standard Normal Distribution

وهو حالة خاصة من التوزيع الطبيعي العام تتميز بما يلي :

1 . أن مركزه ينطبق على مبدأ الإحداثيات، أي أن متوسطه يساوي الصفر $\bar{x} = 0$.

2 . أن تباينه σ^2 يساوي الواحد $\sigma^2 = 1$. ولذلك يرمز له بالرمز N(0,1)

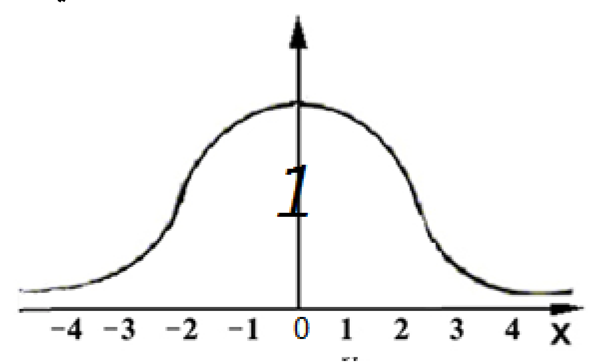
3 . أن المساحة تحته تساوي الواحد.

4 . أنه متناظر بالنسبة للمحور العمودي.

وتميزاً له عن التوزيع الطبيعي العام نرمز لمتحوله المعياري بالرمز Z . وبذلك فإن معادلته تأخذ الشكل التالي:

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-Z^2}{2}} \qquad -\infty < z < \infty \qquad (18-3)$$

وهو يرسم على المحاور الإحداثية منحنياً متناظراً بالنسبة للمحور الشاقولي كما يلي:



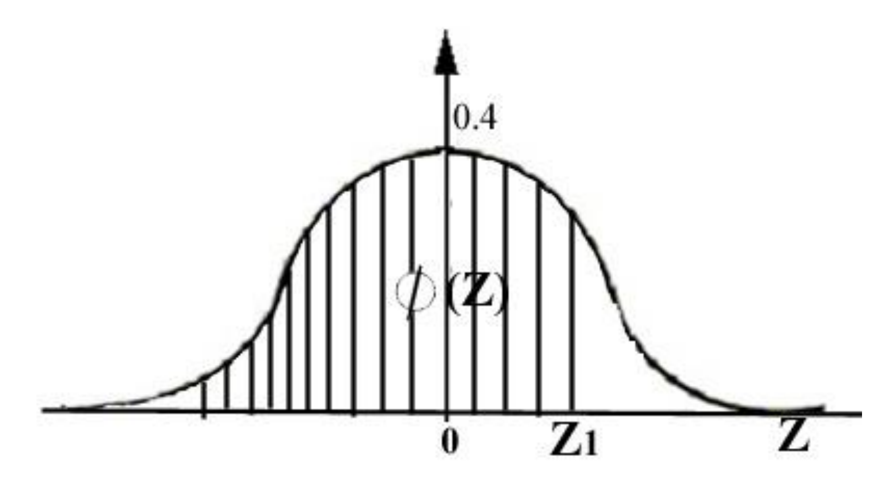
الشكل رقم (3-5): منحنى التوزيع الطبيعي المعياري

وتخضع لهذا التوزيع العديد من المتحولات العشوائية العادية والمتحولات الاصطناعية. ويستخدم هذا التوزيع في حساب الاحتمالات الطبيعية الخاصة والعامة.وسنشرح ذلك كما يلي:

أولاً: كيفية حساب الاحتمالات الخاصة الخاضعة للتوزيع الطبيعي المعياري نفسه:

لحساب هذه الاحتمالات تم إعداد جدول خاص بقيم الاحتمالات التراكمية المختلفة من $(-\infty)$ وحتى قيمة معينةz_1 للمتحول Z، وهذه الاحتمالات هي عبارة عن المساحة تحت المنحنى من $(-\infty)$ حتى القيمة المتداولة Z_1 . وهي المساحة المخططة على الشكل (3-6), وسنرمز لها بـ $\Phi(Z)$، وبذلك يمكننا أن نحسب الاحتمال التراكمي بواسطة قيم $\Phi(Z)$، والذي يسمى بتابع التوزيع الطبيعي المعياري كما يلي:

$$P(Z < z_1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_1} e^{\frac{-z^2}{2}} dz \quad = \Phi(Z_1) \qquad (19-3)$$



الشكل رقم (3−6): تابع التوزيع الطبيعي المعياري $\Phi(Z_1)$

وباختصار نكتب العلاقة السابقة على الشكل التالي:

$$P(Z<Z1)=\Phi(Z_1) \qquad (19-3)$$

وهنا نشير إلى أن قيم تابع التوزيع االطبيعي المعياري $\Phi(Z_1)$ المقابلة لجميع قيم Z الموجبة, محسوبة في الجدول (3−1), ويمكن استخدامه في حساب جميع أنواع الاحتمالات الطبيعية, كما سنرى في الأمثلة القادمة .

مثال(5.2): احسب احتمال أن تكون Z أقل من 1.25, أي $P(Z<1.25)$. ثم احسب احتمالات أن تكون Z أقل من قيم مختارة أخرى.

الحل: 1 . نلاحظ أن قيمة Z المتداولة تساوي $z_1=1.25$.

2 . نبحث في الجدول(3−1) اللاحق وفي عمود Z عن العدد 1.2 ، ثم نسير في سطره حتى نتقاطع مع العمود الموافق لـ 0.05 , فنحصل منه على الرقم 0.8944 المقابل للقيمة المتداولة 1.25. وإن هذا الرقم هو الاحتمال التراكمي المقابل لقيم المتحول Z التي هي أصغر من 1.25.

ونكتب ذلك على الشكل التالي:

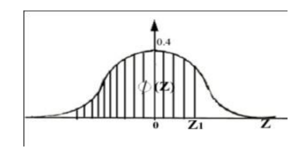
$$P(Z<1.25)=\Phi(1.25)=0.8944$$

3. وكذلك نجد على سبيل المثال أن:

$$P(Z<2.71)=\Phi(2.71)=0.9966$$
$$P(Z<0.57)=\Phi(0.57)=0.7157$$
$$P(Z<0)=\Phi(0)=0.500$$
$$P(Z<1)=\Phi(1)=0.8413$$
$$P(Z<2)=\Phi(2)=0.9772$$

جدول (3−1): قيم تابع التوزيع المعياري المعرف على المجال $]-\infty, z_1[$ **بالعلاقة:**

$$P(Z < z_1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_1} e^{\frac{-z^2}{2}} .dz \qquad = \Phi(z_1)$$



Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983

وهنا نلاحظ من العلاقة (3−19) أن قيم المتحول Z يمكن أن تكون موجبة أو سالبة، ولكن الجدول (3−1) لا يتضمن غير القيم الموجبة لـZ. فلماذا ذلك؟. **الجواب:** لأن التوزيع المعياري متناظر، وهذا ما يمكننا من حساب الاحتمالات المقابلة لقيم Z السالبة من نفس جدول قيم الاحتمالات المقابلة لقيم Z الموجبة, أي من نفس الجدول (3−1) ثم طرحها من العدد واحد.

وكمثال على ذلك نأخذ القيمة (-1.37) كقيمة متداولة, فنجد أن ذلك الاحتمال يساوي:

$$P(Z<-1.37)=1-P(Z<1.37)=1-\Phi(1.37)\quad=1-0.9147=0.0853$$

ويمكن التعبير رياضياً عن ذلك بواسطة العلاقة التالية:

$$P\left(Z<-z_1\right)=1-P(Z<z_1)$$

$$\Phi\left(-Z\right)=1-\Phi\left(Z\right)\qquad(20-3)$$

ولحساب الاحتمال المقابل لمجال جزئي مثل $[\mathbf{z_1},\mathbf{z_2}]$ نقوم بطرح الاحتمال التراكمي المقابل لقيمة $\mathbf{z_1}$ الصغيرة من الاحتمال التراكمي المقابل لقيمة $\mathbf{z_2}$ الكبيرة.

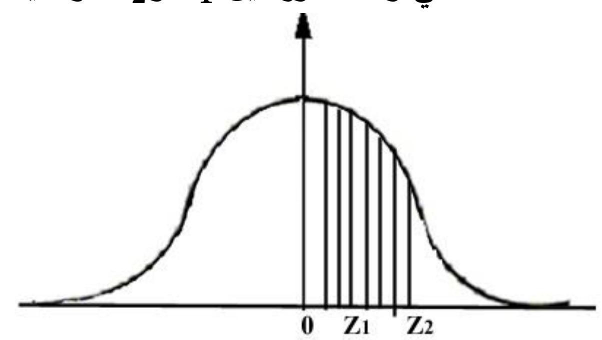
وكمثال على ذلك نأخذ المجال التالي]1.28, 2.37[فنجد أن:

$$\mathbf{P(1.28<Z<2.37)=P(Z<2.37)-P(Z<1.28)}$$
$$\mathbf{=\phi(2.37)-\phi(1.28)\quad=0.9911-0.8997\;=\;\;0.0914}$$

وبصورة عامة يمكنا كتابة ذلك على شكل علاقة عامة كما يلي:

$$\mathbf{P(z_1<Z<z_2)=\phi(z_2)-\phi(z_1)}\qquad\mathbf{(21-3)}$$

وهذا الاحتمال يمثل بالمساحة تحت المنحني والمحصورة بين Z_1 و$\mathbf{Z_2}$، والمبينة كما يلي :



شكل (3−7):الاحتمال على المجال $[\mathbf{z_1},\mathbf{z_2}]$.

ولهذا قمنا بتقدم جدول تفصيلي بقيم تابع التوزيع الطبيعي المعياري $\mathbf{\phi(z)}$ المقابل للقيم الموجبة لـ z , والذي يمكن استخدامه لحساب جميع الاحتمالات المختلفة بتطبيق العلاقات السابقة.

ثانياً: حساب الاحتمالات للمتحولات X الخاضعة للتوزيع الطبيعي العام:

لحساب تلك الاحتمالات ننطلق من نظرية هامة وبسيطة وهي:

نظرية: إذا كان X خاضعاً للتوزيع الطبيعي العام(3−17)، الذي متوسطه $\bar{\mathbf{x}}$ وانحرافه المعياري σ، فإن المتحول الاصطناعي Z التالي: $\mathbf{Z=\frac{X-\bar{x}}{\sigma}}$ يخضع للتوزيع الطبيعي المعياري $N(0,1)$.

لذلك تُسمى التحويلة $\mathbf{Z=\frac{X-\bar{x}}{\sigma}}$ بالدرجة المعيارية أو المتحول المعياري، ويُستفاد من هذه التحويلة في حساب الاحتمالات العامة، حيث يتم تحويل كل متحول طبيعي عام X إلى المتحول المعياري Z , وذلك بحساب الدرجة المعيارية المقابلة له، ثم استخدام الجدول (3−1) لتابع التوزيع المعياري في حساب الاحتمالات المطلوبة.

وبشكلٍ عام نجد أنه تكون لدينا عدة حالات لحساب الاحتمالات العامة، هي:

1 . حساب الاحتمال العام الذي من الشكل $P(x \le x_1)$.

لحساب هذا الاحتمال نطرح من طرفي المتراجحة $\bar{x}$ ثم نقسّم الناتج على σ فنحصل على ما يلي:

$$P(x \le x_1) = P\left[\frac{x-\bar{x}}{\sigma} \le \frac{x_1-\bar{x}}{\sigma}\right]$$

وباستخدام رمز التحويلة المعيارية السابقة نجد أن:

$$Z = \frac{x-\bar{x}}{\sigma} \quad , \qquad Z_1 = \frac{x_1-\bar{x}}{\sigma}$$

وبناء على ذلك نجد أن الاحتمال السابق:

$$P(x \le x_1) = P(Z \le Z_1) = \phi(Z_1)$$

أي أن حساب الاحتمال السابق يعطى بالعلاقة التالية:

$$P(x \le x_1) = \phi\left(\frac{x_1-\bar{x}}{\sigma}\right) \qquad (22-3)$$

مثال(3−6): لنفترض أن X هي درجة الطالب في الامتحان وأن متوسطها $\bar{x} = 65$ وأن انحرافها المعياري 14= σ، فأوجد احتمال أن تكون درجة الطالب أقل من **70**، وذلك بفرض أن X يخضع للتوزيع الطبيعي العام.

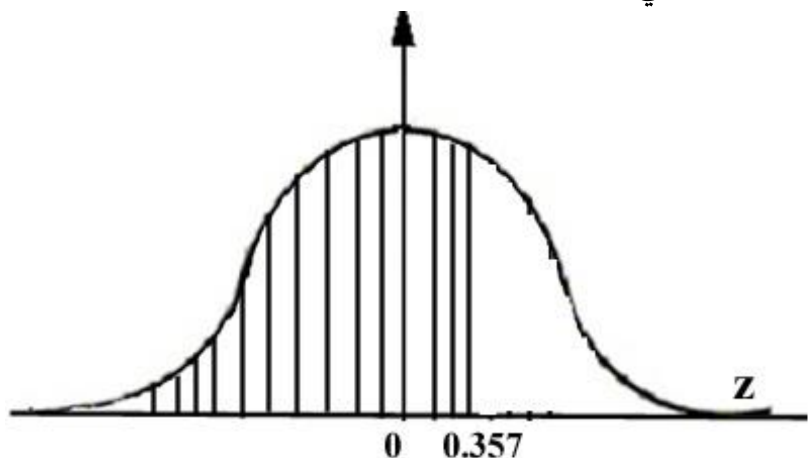
الحل: نلاحظ أن X يخضع للتوزيع الطبيعي العام وليس للتوزيع المعياري. ولإيجاد الاحتمال المطلوب نقوم بتحويل X إلى Z وحساب الدرجة المعيارية المقابلة للقيمة المطلوبة $x_1 = 70$, فنجد أن:

$$Z_1 = \frac{x_1-\bar{x}}{\sigma} = \frac{70-65}{14} = \frac{5}{14} = 0.357$$

ومن التوزيع الطبيعي المعياري نجد أن:

$$P(x < 70) = P(z < 0.357) = \Phi(0.357) = 0.6406$$

وهو الاحتمال المطلوب, أي أن احتمال أن تكون درجة الطالب أقل من 70 يساوي 0.6406 , كما هو مبين على الشكل المخطط التالي:



الشكل (6−8): الاحتمال المطلوب $P(Z < 0.357)$

وهكذا يمكننا أن نحسب جميع الاحتمالات المختلفة الأخرى المشابهة لذلك.

2 . حساب الاحتمال العام في مجال محدد $[x_1, x_2]$ من الشكل $P(x_1 \leq X \leq x_2)$:

لحساب هذا الاحتمال نستخدم الأسلوب نفسه, فنجد أن:

$$P(x_1 \leq X \leq x_2) = P\left[\frac{x_1 - \bar{x}}{\sigma} \leq \frac{X - \bar{x}}{\sigma} \leq \frac{x_2 - \bar{x}}{\sigma}\right]$$

$$P(x_1 \leq X \leq x_2) = P[z_1 \leq Z \leq z_2] = \phi(z_2) - (z_1)$$

$$P(x_1 \leq X \leq x_2) = \phi\left(\frac{x_2 - \bar{x}}{\sigma}\right) - \phi\left(\frac{x_1 - \bar{x}}{\sigma}\right) \qquad (23 - 3)$$

مثال(3−7): لنأخذ معطيات المثال السابق ونحسب احتمال أن تكون درجة الطالب واقعة في المجال [80، 55]، لذلك نقوم في البداية بحساب الدرجة المعيارية المقابلة لكل من طرفي المجال, فنجد أن:

$$Z_1 = \frac{x_1 - \bar{x}}{\sigma} = \frac{55 - 65}{14} = -0.714$$

$$Z_2 = \frac{x_2 - \bar{x}}{\sigma} = \frac{80 - 65}{14} = 1.07$$

ونحصل على الاحتمال المطلوب باستخدام الدرجتين المعياريتين لطرفي المجال, فنجد أن:

$$P(x_1 \leq X \leq x_2) = \phi(z_2) - \phi(z_1)$$

ومن جدول تابع التوزيع المعياري نجد أن:

$$P(55 < X < 80) = P(-0.714 < Z < 1.074) = \Phi(1.07) - \Phi(-0.714)$$

$$= \Phi(1.07) - [1 - \Phi(0.714)] = 0.8577 - [1 - 0.7612] = 0.6189$$

وهو يمثل المساحة الواقعة فوق المجال [1.0747 , 0.714−] للتوزيع الطبيعي المعياري, كما هو مبين في الشكل (3−9).

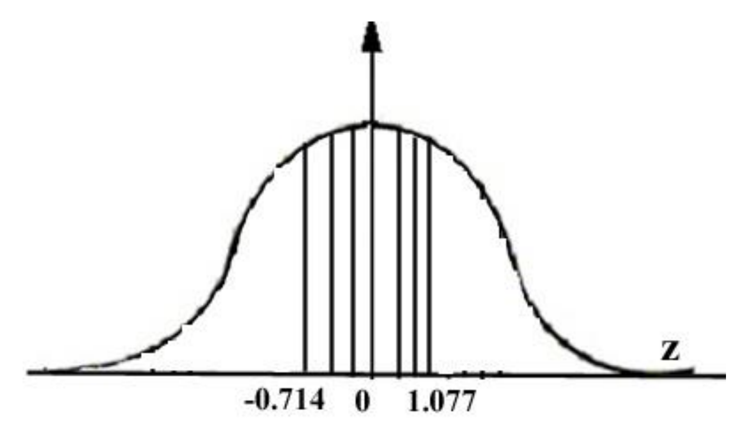
وبصورة عامة يمكننا أن نطبق العلاقة(3−15) السابقة مباشرةً لحساب الاحتمالات العامة واستخدام الصيغة التالية:

$$P(x_1 \leq X \leq x_2) = \phi\left(\frac{x_2 - \bar{x}}{\sigma}\right) - \phi\left(\frac{x_1 - \bar{x}}{\sigma}\right)$$

حيث نجد أن:

$$\begin{aligned} P(55 \leq X \leq 80) &= \phi\left(\frac{80 - 65}{14}\right) - \phi\left(\frac{55 - 65}{14}\right) \\ &= \phi(1.071) - \phi(-0.714) \\ &= \phi(1.071) - [1 - \phi(-0.714)] \\ &= 0.8577 - (1 - 0.7612) \end{aligned}$$

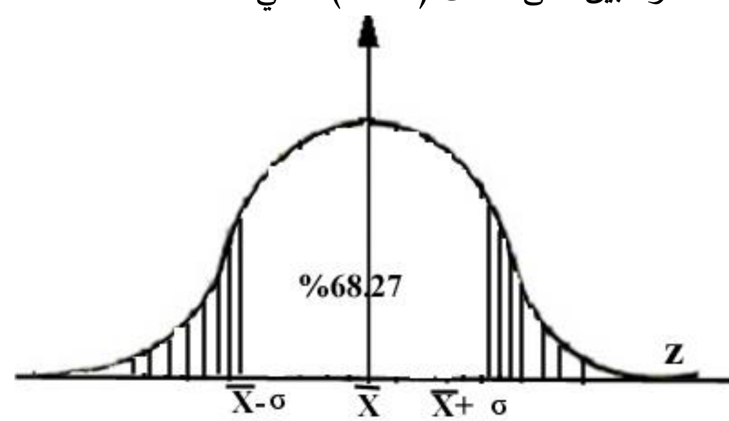
$$P(55 \leq X \leq 80) = 0.6189$$



الشكل (3−9): الاحتمال المطلوب

3 . حساب الاحتمالات المقابلة لمجالات الثقة الأساسية، وهي:

- **الاحتمال المقابل لمجال الثقة الأول:** وهو الاحتمال المقابل للمجال الذي مركزه المتوسط $\bar{x}$ وطرفاه محددين بالعددين $\bar{x}+\sigma$ و$\bar{x}-\sigma$، أي محددين بجمع وطرح الانحراف المعياري من المتوسط $\bar{x}$. كما هو مبيّن على الشكل (2−10) التالي:



الشكل (3−10): مجال الثقة الأول

إن الاحتمال المطلوب هو: $\mathbf{P(\bar{x}-\sigma \le X \le \bar{x}+\sigma)}$

لحساب هذا الاحتمال نقوم بإيجاد الدرجة المعيارية المقابلة لكل طرفي المجال كما يلي:

$$\mathbf{z_1 = \frac{(\bar{x}-\sigma)-\bar{x}}{\sigma} = -1}$$

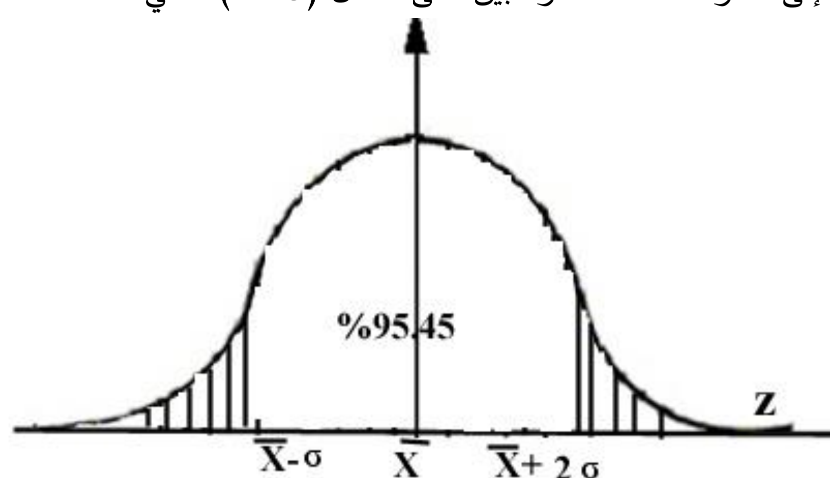
$$\mathbf{z_2 = \frac{(\bar{x}+\sigma)-\bar{x}}{\sigma} = +1}$$

ومن الجدول المعياري نجد أن الاحتمال المطلوب يساوي:

$$\begin{aligned} P(\bar{x}-\sigma \le X \le \bar{x}+\sigma) &= \phi(z_2)-\phi(z_1) \\ &= \phi(1)-\phi(-1) \qquad (24-3) \\ &= 0.8413-(1-0.8413) = 0.6826 \end{aligned}$$

وهو يساوي المساحة غير المخططة فوق المجال الأول والمبنية على الشكل (3−10)، وهذا يعني أن احتمال أن يقع X في ذلك المجال يساوي 0.6826، أي أن 68% فقط من قيم X تقع في ذلك المجال.

• **الاحتمال المقابل لمجال الثقة الثاني:** وهو الاحتمال المقابل للمجال الذي مركزه المتوسط $\bar{\mathbf{x}}$ وطرفاه محددين بالعددين $\bar{\mathbf{x}}+\mathbf{2\sigma}$ و$\bar{\mathbf{x}}-\mathbf{2\sigma}$، أي محددين بجمع وطرح ضعفي الانحراف المعياري إلى المتوسط $\bar{\mathbf{x}}$. كما هو مبيّن على الشكل (3−11) التالي:



الشكل (3−11): مجال الثقة الثاني

إن الاحتمال المطلوب هو: $\mathbf{P(\bar{x}-2\sigma \le X \le \bar{x}+2\sigma)}$

لحساب هذا الاحتمال نقوم بإيجاد الدرجة المعيارية المقابلة لكل من طرفي المجال كما يلي:

$$\mathbf{z_1=\frac{(\bar{x}-2\sigma)-\bar{x}}{\sigma}=-2}$$

$$\mathbf{z_2=\frac{(\bar{x}+2\sigma)-\bar{x}}{\sigma}=+2}$$

وبذلك نجد أن الاحتمال المطلوب يساوي:

$$P(\bar{x}-2\sigma \le X \le \bar{x}+2\sigma)=\phi(z_2)-\phi(z_1)=\phi(2)-\phi(-2)$$

ومن جدول تابع التوزيع المعياري نجد أن الاحتمال المطلوب يساوي:

$$P(\bar{x}-2\sigma \le X \le \bar{x}+2\sigma)=0.9772-(1-0.9772)$$

$$P(\bar{x}-2\sigma \le X \le \bar{x}+2\sigma)=0.9544 \qquad \mathbf{(25-3)}$$

وهو يساوي المساحة غير المخططة فوق المجال الثاني والمبنية على الشكل (3−11) أعلاه، وهذا يعني أن احتمال أن يقع X في المجال الثاني يساوي 0.9544، أي أن حوالي 95% فقط من قيمه الممكنة تقع في ذلك المجال.

• **الاحتمال المقابل لمجال الثقة الثالث: وهو** $\mathbf{P(\bar{x}-3\sigma \le X \le \bar{x}+3\sigma)}$

بطريقة مشابهة نجد أن احتمال وقوع X في المجال الثالث للثقة يساوي:

$$P(\bar{x}-3\sigma \le X \le \bar{x}+3\sigma)=0.9973 \qquad (26-3)$$

ملاحظة: هناك تطبيقات هامة كثيرة لهذه المجالات ولاحتمالاتها المقابلة لذلك يجب على الطلاب إعطاءها اهتماماً خاصاً.

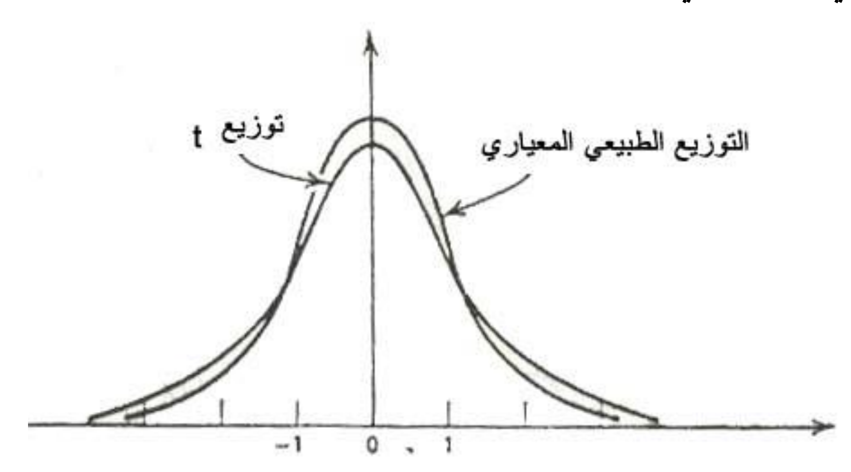
3−2−3 : توزيع ستودينت (توزيع t):Student’s Distribution

وهو يستخدم بدلاً عن التوزيع الطبيعي المعياري, عندما يكون حجم العينة صغيراً(أي عندما يكون $n \leq 30$), وإن الشكل العام لمنحني هذا التوزيع يشبه منحني التوزيع الطبيعي المعياري، فهو متناظر بالنسبة لمبدأ الإحداثيات (أي أن مركزه صفر) ولكن انحرافه المعياري يختلف عن الواحد. وأن صيغة معادلته الرياضية هي:

$$f(t) = C\left(1+\frac{t^2}{k}\right)^{-\frac{k+1}{2}} \qquad (26-3)$$

حيث أن: $-\infty < \mathbf{t} < +\infty$ ويُسمى العدد k بدرجة الحرية، و C ثابت عددي موجب ومعين .

ويأخذ شكله البياني الشكل الآتي:



الشكل رقم (3−12): منحني توزيع ستودينت

وغالباً ما يرمز للمتغير العشوائي الخاضع لتوزيع ستودينت بالرمز t، ويُستخدم هذا التوزيع في اختبارات الفرضيات المختلفة عندما تكون العينات صغيرة الحجم (حجم العينة n أقل من 30). وعندما يقوم الباحث بحساب قيمة مؤشر الاختبار للفرض الذي يختبره, فإنه يحصل على قيمة عددية نرمز لها بـ t. ثم يقوم بمقارنة هذه القيمة مع القيمة الحرجة $\mathbf{t_0}$ المقابلة لمستوى الدلالة ولدرجة الحرية، (ونحصل على قيمة $\mathbf{t_0}$ من جداول توزيع t الملحقة)، فإذا كانت القيمة المطلقة لـ t المحسوبة أصغر من القيمة الحرجة $\mathbf{t_0}$، نقبل فرضية العدم الموضوعة، وإذا كان العكس نرفض فرضية العدم الموضوعة ونقول بعكسها، لأننا لا نملك دليلاً كافياً على إثباتها. وسنتطرق إلى هذا الموضوع في اختبار الفرضيات الواردة في الفصل السادس .

4−2−3 : توزيع كاي مربع χ^2:χ^2 Distribution

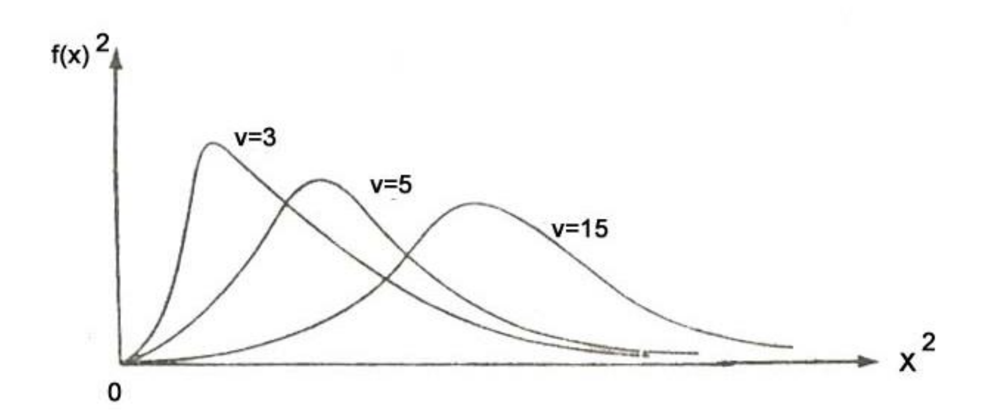
وهو توزيع احتمالي خاص يُستخدم لاختبار مدى تطابق المعطيات الميدانية أو التجريبية مع المعطيات الفرضية أو الضابطة. ولتطبيقه تُستخدم علاقة رياضية معينة لحساب قيمة مؤشر الاختبار $\boldsymbol{\chi^2}$ ثم مقارنتها مع القيمة الحرجة $\boldsymbol{\chi_0^2}$ المقابلة لمستوى الدلالة $\boldsymbol{\alpha}$ ودرجة الحرية المحددة k، ويتم الحصول على $\boldsymbol{\chi_0^2}$ من جداول خاصة بالتوزيع $\boldsymbol{\chi^2}$. فإذا كانت قيمة $\boldsymbol{\chi^2}$ المحسوبة أصغر من $\boldsymbol{\chi_0^2}$ الحرجة نقبل

فرضية العدم الموضوعة، كما هو موضح على الشكل التالي، وفي حالة العكس نرفض تلك الفرضية لعدم وجود دليل كافٍ على إثباتها.

وإن الصيغة الرياضية لتوزيع χ^2 هي كما يلي:

$$f(x^2) = C \cdot (\chi^2)^{\frac{k}{2}-1} e^{-\frac{\chi^2}{2}} \quad : \chi^2 > 0 \qquad (28-3)$$

حيث أن: k درجة الحرية، وC ثابت عددي موجب ومعين . ويأخذ شكله البياني الشكل الآتي:



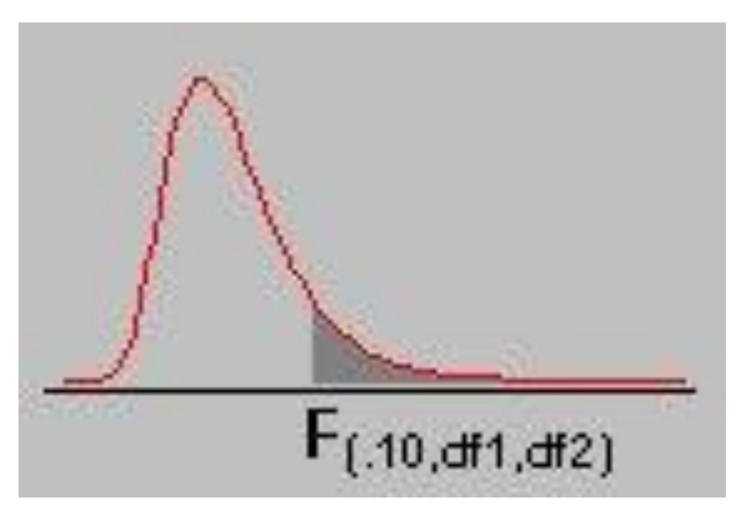
الشكل (4–13): منحني توزيعχ^2 (v درجة الحرية)

5–2–3 : توزيع فيشر أو توزيع F: Fisher's Distribution

وهو توزيع احتمالي خاص ومعقد, يستخدم لاختبارات التجانس بين المجتمعات ولتحليل التباين لمعطيات العينات، ولتطبيقه تُستخدم علاقة رياضية معينة لحساب قيمة مؤشر الاختبار F، ثم مقارنة هذه القيمة المحسوبة مع القيمة الحرجة $\mathbf{F_0}$ المقابلة لمستوى الدلالة $\boldsymbol{\alpha}$ ولدرجتي الحرية في العينتين. فإذا كانت قيمة F المحسوبة أصغر من $\mathbf{F_0}$ الحرجة نقبل فرضية العدم الموضوعة، وفي حالة العكس نرفض تلك الفرضية لعدم وجود دليل كافٍ على إثباتها، كما هو موضح على الشكل التالي. وتكون معادلته الرياضية كما يلي:

$$f(F) = C \cdot \frac{F^{\frac{k}{2}-1}}{k_2 + k_1 F^{\frac{k_2+k_1}{2}}} \quad : F > 0 \qquad (29-3)$$

حيث أن: k_1 درجة الحرية للتباين الكبير الذي في البسط ,وk_2 درجة الحرية للتباين الثاني الذي في المقام، وC ثابت عددي موجب ومعين .وإنه يرسم الشكل البياني الآتي:



الشكل (3−14): منحني توزيع F

6-2-3 : تقريب التوزيع الثنائي بواسطة التوزيع الطبيعي المعياري :

عندما يكون عدد التجارب في التوزيع الثنائي كبيراً نسبياً وتكون قيمة احتمال التحقق العام ***p*** غير صغيرة فإن عملية حساب احتمالات التوزيع الثنائي تصبح معقدة جداً (يسبب حساب C^k_n).

وفي مثل هذه الحالات يمكننا حساب قيم تقريبية للتوزيع الثنائي من خلال التوزيع الطبيعي المعياري وذلك بتطبيق العلاقة (6−23) على مجال حول القيمة المطلوبة ***k*** نصف طوله يساوي $\frac{1}{2}$ وحدة، فنجد أن :

$$p(X=k)=C_n^k\,p^k\,q^{n-k}=p\left(k-\frac{1}{2}\le X\le k+\frac{1}{2}\right)$$

$$=\ \Phi\left[\frac{\left(k+\frac{1}{2}\right)-\ n.p}{\sqrt{n.p.q}}\right]-\Phi\left[\frac{\left(k-\frac{1}{2}\right)-n\,p}{\sqrt{n.p.q}}\right]\qquad(30-3)$$

ولكن يشترط في هذه الحالة أن يتحقق الشرطان التاليان معاً: $5\le np$ و $5\le nq$

مثال(3−8): احسب احتمال أن ينجح 155 طالباً تماماً من أصل 200 طالباً، إذا علمت بأن نسبة النجاح كانت **p = 0.75** .

الحل : لنرمز لعدد الناجحين بـ **X** ، لحساب هذا الاحتمال بدقة يجب أن نطبق التوزيع الثنائي كما يلي :

$$p(X=155)=C_{200}^{155}\,p^{155}.q^5=??$$

وبما أن حساب هذا الاحتمال صعب جداً لذلك نلجأ إلى حساب قيمة تقريبية له من العلاقة (3− 30) فنجد أن الشرطين **n q = 150 > 5** و **n p = 50 > 5** *محققان*، وبذلك يكون لدينا:

$$p(X=155)=\Phi\left[\frac{155.5-200\,(0.75)}{\sqrt{200\,(0.75)(0.25)}}\right]-\Phi\left[\frac{154.5-200(0.75)}{\sqrt{200(0.75)(0.25)}}\right]$$

$$= \Phi\left[\frac{+5.5}{6.124}\right] - \Phi\left[\frac{+4.5}{6.124}\right]$$
$$= \Phi(0.898) - \Phi(0.735)$$
$$= 0.8159 - 0.7704 = 0.0455$$

وبشكل عام إذا أردنا حساب احتمال أن يقع ***x*** الخاضع للتوزيع الثنائي في مجال محدد فنطبق العلاقة :

$$p(k_1 \leq X \leq\ k_2) \quad = p\left[k_1 - \frac{1}{2} \leq X \leq k_2 + \frac{1}{2}\right] \qquad (31-3)$$

$$p(k_1 \leq X \leq k_2) = \Phi\left[\frac{\left(k_2 + \frac{1}{2}\right) - np}{\sqrt{n.p.q}}\right] - \Phi\left[\frac{\left(k_1 - \frac{1}{2}\right) - np}{\sqrt{n.p.q}}\right]$$

وكمثال على ذلك نحسب احتمال أن يكون عدد الناجحين في المقرر السابق محصوراً بين ***140*** طالباً و 170 طالباً. فنجد أن:

$$p(140 \leq X \leq 170) = \Phi\left[\frac{\left(170+\frac{1}{2}\right)-150}{\sqrt{200*0.75*0.25}}\right] - \Phi\left[\frac{\left(140-\frac{1}{2}\right)-150}{\sqrt{200*0.75*0.25}}\right]$$
$$= \Phi[3.35] - \Phi[-1.71]$$
$$= \Phi[3.35] - [1 - \Phi(1.71)]$$
$$= 0.9996 - [1 - 0.9564] = 0.956$$

تمرينات

1 . إذا كان Z خاضعاً للتوزيع الطبيعي المعياري فاحسب الاحتمالات الآتية:

، $P(-1 \leq Z \leq +1)$, $P(-2 \leq Z \leq +2)$

، $P(-2.58 \leq Z \leq +2.58)$, $P(-3 \leq Z \leq +3)$

2 . إذا كانت X قيمة المبيعات اليومية في إحدى الصيدليات تخضع للتوزيع الطبيعي وكان متوسطها $\overline{X} = 3000$ وتباينها $\sigma^2 = 14900$ والمطلوب:

. تحديد الصيغة الرياضية لقانون التوزع الاحتمالي لـ X.

. حساب احتمال أن تتجاوز قيمة المبيعات اليومية 4000 ل.س، أي $P(X > 4000)$

. حساب الاحتمال التالي: $P(2000 \leq X \leq 4000)$

3 . إذا كان وزن الطفل عند الولادة X يخضع للتوزيع الطبيعي العام الذي متوسطه $\bar{\mathbf{x}} = \mathbf{3}$ kg. وتباينه $\boldsymbol{\sigma^2} = \mathbf{0.64}$, فالمطلوب:

. تحديد الصيغة الرياضية لقانون التوزيع الاحتمالي لـ X.

. حساب الاحتمالات التالية:

$P(2 \leq X \leq 4)$ ، $P(X \geq 3)$ ، $P(X \leq 2)$ ، $P(X \geq 4)$

4 . إذا كان X هو مقدار الدخل اليومي لعيادة أحد الأطباء وكان خاضعاً للتوزيع الطبيعي الذي متوسطه $\bar{x} = 4000$ وتباينه $\sigma^2 = 490000$, فالمطلوب:

. تحديد الصيغة الرياضية لقانون التوزيع الاحتمالي للمتحول لـ X.

. حساب الاحتمالات التالية:

$P(X \leq 5000)$ ، $P(X \leq 5000)$ ، $P(3000 \leq X \leq 5000)$

$P(2000 \leq X \leq 5000)$

5 . إذا كان X هو طول الشخص البالغ وكان خاضعاً للتوزيع الطبيعي العام الذي متوسطه $\bar{x} = 170$ وتباينه $\sigma^2 = 25$, فالمطلوب:

. تحديد الصيغة الرياضية للطول لـ X.

. حساب الاحتمالات التالية:

، $P(X \geq 180)$ ، $P(160 \leq X \leq 180)$ ، $P(X \leq 160)$

$P(\bar{x} - 2\sigma \leq X \leq \bar{x} + 2\sigma)$

الفصل الرابع
العينات ومسائل التقدير

تمهيد:

تعتبر طريقة العينات من الطرائق الإحصائية المستخدمة في عمليات جمع المعلومات الإحصائية وتحليلها، وذلك لأنها توفر كثيراً من المال والجهد والوقت في الحصول على المعلومات اللازمة.

وتعد طريقة العينات في كثير من الأحيان الأسلوب الإحصائي الوحيد الذي يمكن تطبيقه في الحصول على المعلومات، كما في حالة تحليل الدم أو تحليل التربة، أو تحليل المياه، أو مراقبة الإنتاج، أو تقدير صلاحية الأغذية أو الأدوية...إلخ.

4-1: أنواع العينات:

تصنّف العينات إلى عدة أنواع حسب تصميم عمليات المعاينة , والذي يعتمد على طبيعة المجتمع المدروس وعلى هدف الدراسة... وأنواعها هي:

1 . **المعاينة العشوائية البسيطة:** وهي التي يتم فيها سحب عينة عشوائية من عناصر المجتمع المدروس وباحتمالات متساوية, وتطبق على المجتمعات المتجانسة.

2 . **المعاينة العشوائية الطبقية:** وتطبق على المجتمعات غير المتجانسة , وفيها يتم تقسيم المجتمع المدروس غير المتجانس إلى طبقات متجانسة، ثم تطبيق المعاينة البسيطة على كل طبقة على حده، ثم يتم تركيب النتائج وإجراء الحسابات اللازمة.

3 . **المعاينة العنقودية:** وتطبق على المجتمعات, التي لها تركيبة عنقودية, وفيها يتم سحب العينات العنقودية على مرحلتين أو أكثر: تُسحب عينة فروع من الجذع ثم تسحب عينات عناصر من الفروع المسحوبة (مثل سحب عينة من المدن، ثم عينات من الأسر من كل مدينة مسحوبة).

4 . **المعاينة المنتظمة :** وتطبق على المجتمعات المتحركة, وفيها يتم سحب عناصر العينة بطريقة سلسلة عددية منتظمة, كأن نسحب عنصراً بعد مرور كل k عنصراً.

اما خطوات البحث المتبعة في طريقة العينات فهي نفسها الخطوات المتبعة في حالة المسح الشامل, غير أن طريقة العينات تحتاج إلى خطوة إضافية وهامة هي خطوة سحب العينة، وهو ما سنعالجه في البند التالي.

- **كيفية سحب العينة العشوائية البسيطة:**

يُشترط في سحب العينة أن يكون سحباً عشوائياً ولذلك نقوم بما يلي:

1 . نحصر عناصر المجتمع المدروس ونحدد عددها N.

2 . نعطي كل عنصر رقماً متسلسلاً من 1 حتى N.

3 . نضع نسخة عن أرقام هذه العناصر في صندوق خاص ونخلطها جيداً.
4. نقوم بسحب الأرقام من الصندوق بطريقة عشوائية ونخلط الأرقام قبل كل عملية سحب.
5. نعتبر الأرقام المسحوبة في الخطوة السابقة هي أرقام عناصر المجتمع التي ستدخل في العينة، ونقوم بسحبها من المجتمع لدراستها واستخلاص المعلومات اللازمة منها.
أما السحب بحد ذاته فيمكن أن يكون على شكلين هما:
. السحب مع الإعادة: وفيه يُعاد الرقم المسحوب إلى الصندوق ليتعرض للسحب مرةً ثانية أو ثالثة ...الخ.
. السحب بدون إعادة، وفيه لا يُعاد الرقم المسحوب إلى الصندوق وبذلك لا يتعرض للسحب مرةً ثانية.
ويمكننا سحب الأرقام عشوائياً بواسطة استخدام عجلات الرهان، أو بواسطة جداول الأرقام العشوائية، أو بواسطة علاقات رياضية مبرمجة على الآلات الحاسبة أو الحواسيب العادية.
إن عملية السحب العشوائي لعناصر العينة تضمن لنا عدم التحيّز في الانتقاء، كما تؤمّن تمثيلاً جيداً للمجتمع المدروس, وتضمن لنا الحصول على تقديرات جيدة لمعالم ذلك المجتمع.
وقبل أن نبحث في قضايا التقدير نورد بعض التعاريف المتعلقة بالمجتمع والعينة والمؤشرات المستخدمة في تقدير معالم المجتمع.

2-4 : تعريف المجتمع الإحصائي والعينة:

- **المجتمع الإحصائي ومعالمه** : هو جملة العناصر التي تكون مستهدفة بالدراسة , وسنرمز لعدد تلك العناصر بـ N، وللخاصة المستهدفة بالدراسة فيه بـ Y, ولقيمها المختلفة بالرموز.

$$Y:\ y_1, y_2, y_3, y_4 \ldots\ldots\ldots, y_i, \ldots\ldots, y_n$$

ونرمز لمتوسط هذه القيم في المجتمع بالرمز $\bar{y}$ والذي يساوي:

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \qquad (1-4)$$

ونرمز لتباين تلك القيم بالرمز σ^2 وهو يساوي:

$$\sigma^2 = \frac{1}{N}\cdot\sum_{i=1}^{N}(y_i - \bar{y})^2 \qquad (2-4)$$

ولنسبة وجود تلك الخاصة في المجتمع بالرمز R وتساوي:

$$R = \frac{M}{N} \qquad (3-4)$$

حيث M هو عدد المتصفين بتلك الخاصة في المجتمع.
وبما أن هذه القيم ومتوسطها وتباينها ونسبتها تكون مجهولة في المجتمع، لذلك نسحب عينة من المجتمع بصورة عشوائية، وسنستخدم معلومات تلك العينة لتقدير معالم المجتمع الاساسية: كالمتوسط $\bar{y}$ والتباين σ^2 والنسبة R ... إلخ.

- **العينة ومؤشراتها:** العينة هي جزء من المجتمع مسحوب عشوائياً منه ومؤلف من n عنصراً (ويسمى n بحجم العينة).

وإذا رمزنا للبيانات التي تقدمها عناصر العينة عن نفس الظاهرة المدروسة بالرمز X ولقيمها بالرموز:

$$X : x_1 \ , \ x_2 \ , \ x_3 , \ x_4 \quad , x_n \qquad (4-4)$$

ولنفترض أن متوسطها كان يساوي:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (5-4)$$

وأن تباينها المصحح كان يساوي:

$$s^2 = \frac{1}{n-1}\cdot\sum\left(x_i - \bar{x}\right)^2 \qquad (6-4)$$

وأن نسبة الخاصة المدروسة في العينة كانت تساوي:

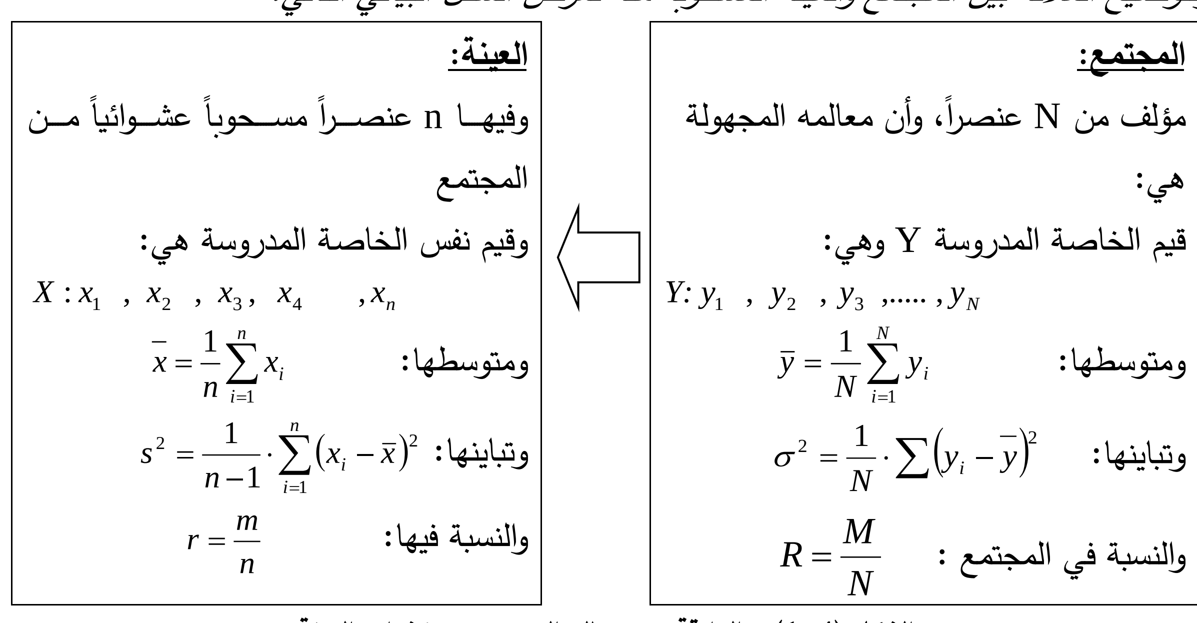
$$r = \frac{m}{n} \qquad (7-4)$$

وسنقوم باستخدام مؤشرات هذه العينة لتقدير معالم المجتمع. ولكن قبل ذلك نذكر بمعايير جودة التقدير وهي:

- **معايير جودة التقدير:** نظراً لأن عدد العينات , التي يمكن سحبها من المجتمع يساوي عدداً كبيراً ، فإن كل عينة تعطينا تقديراً خاصاً للمؤشر المطلوب , لذلك كان لابد من وضع معايير لجودة هذه التقديرات وهي:

1. **عدم التحيز:** وهو يعني أن يكون توقع قيمة التقدير , مأخوذاً على جميع العينات الممكنة, مساوياً لقيمة المؤشر في المجتمع.
2. **التماسك:** وهو يعني أن الصيغة الرياضية المستخدمة في حساب التقدير من العينة تنتهي إلى الصيغة الرياضية المعرفة لحساب المؤشر في المجتمع عندما يصبح حجم العينة كبيراً أو مساوياً لحجم المجتمع.
3. **الفعالية:** وهي تعني أن يكون تباين التقدير أصغر من جميع تباينات التقديرات الأخرى.

ولتوضيح العلاقة بين المجتمع والعينة المسحوبة منه نعرض الشكل البياني التالي:



الشكل (4-1) : العلاقة بين معالم المجتمع و مؤشرات العينة

وسنرى كيف يمكن استخدام مؤشرات هذه العينة لتقدير معالم المجتمع من خلال الفقرات التالية.

4-3 : تقدير متوسط المجتمع:

عندما نريد تقدير متوسط أي مؤشر من معالم المجتمع، فإننا نعتمد بشكل أساسي على معطيات العينة المعلومة. لذلك نرمز لقيم الخاصة المدروسة في العينة التي حجمها n بالرموز التالية:

$$X : x_1 \ , \ x_2 \ , \ x_3 , \ x_4 \quad , x_n$$

وهي قيم معلومة لأنها مأخوذة من العينة المسحوبة أومن استمارات البحث الميداني أو التجريبي. لذلك يمكننا حساب قيمة متوسطها من العلاقة التالية:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{N} x_i \qquad (8-4)$$

ولقد رمزنا لمتوسط ذلك المؤشر في المجتمع بالرمز $\bar{y}$ وهو مقدار مجهول, وعلينا أن نقوم بتقديره من خلال معطيات العينة. ولتقدير ذلك المتوسط $\mathbf{\bar{y}}$ نستخدم متوسط العينة $\mathbf{\bar{x}}$, ونرمز لذلك التقدير بـ $\tilde{\bar{y}}$, ونكتبه على الشكل التالي:

$$\tilde{\bar{y}} = \bar{x} = \frac{\sum_{i=1}^{N} x_i}{n} \qquad (9-4)$$

ويبرهن على أن هذا التقدير هو تقدير غير متحيز وفعال ومتماسك ويعتمد عليه في جميع البحوث والدراسات العلمية.

مثال(4-1):

لتقدير متوسط عمر الفتاة عند الزواج الأول في قرية ما، سحبنا بدون إعادة عينة عشوائية بحجم n=10 من النساء المتزوجات في المجتمع المدروس, فوجدنا أن أعمارهن عند الزواج الأول كانت كما يلي:

X: 20, 23, 26, 21, 25, 20, 24, 23, 25, 23

وبسهولة نجد أن متوسط العمر في هذه العينة يساوي:

$\bar{x} = \frac{230}{10} = 23$ سنة

ولتقدير متوسط العمر في المجتمع نستخدم متوسط العينة ونكتب أن:

$\tilde{\bar{y}} = \bar{x} = 23$ سنة

وهو تقدير غير متحيز وفعال ومتماسك لعمر الفتاة عند الزواج الأول في تلك القرية المذكورة .

4-4: تقدير التباين والانحراف المعياري:

لتقدير تباين أي مؤشر في المجتمع، والذي رمزنا له بـ $\boldsymbol{\sigma^2}$ نستخدم تباين العينة المصحح أو المعدل s^2، وهو التباين المصحح والمعرف بالعلاقة التالية:

$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} \left(x_i - \bar{x} \right)^2 \qquad (10-4)$$

لقد تم البرهان على أن التقدير غير المتحيز لتباين المجتمع $\boldsymbol{\sigma^2}$, هو تباين العينة المصحح s^2، ونكتب ذلك كما يلي:

$$\tilde{\bar{\sigma}} = S^2 = \frac{\sum_{n-1}^{n} \left(x_i - \bar{x} \right)^2}{n-1} \qquad (11-4)$$

وبذلك نجد أن تقدير الانحراف المعياري $\boldsymbol{\sigma}$ للمجتمع يتم بواسطة الانحراف المعياري المصحح للعينة s ، ونكتب ذلك كما يلي:

$$\tilde{\bar{\sigma}} = +\sqrt{s^2} = s \qquad (12-4)$$

ولقد تم ترميز الانحراف المعياري للعينة s على الآلات الحاسبة بالرمز $\boldsymbol{\sigma_{n-1}}$، وذلك انسجاماً مع العلاقة (4-11) السابقة.

مثال(4-2): لنأخذ معطيات العينة الواردة في المثال (4-1) السابق ولنحسب التباين المصحح للعينة من العلاقة المعدلة التالية:

$$s^2 = \frac{\sum_{n-1}^{n} (x_i - \bar{x})^2}{n-1}$$

$$s^2 = \frac{(20-23)^2 + (22-23)^2 + (26-23)^2 + ... + (25-23)^2 + (23-23)^2}{9}$$

$$s^2 = 4.41$$

وبذلك نجد أن تباين المجتمع يقدر بما يلي:

$$\tilde{\bar{\sigma}} = s^2 = 4.41$$

وأن الانحراف المعياري للعمر في المجتمع يقدر بما يلي:

$$\tilde{\sigma} = +\sqrt{s^2} = \sqrt{4.41} = 2.10$$

وهو تقدير غير متحيز وفعال ومتماسك للانحراف المعياري للعمر المذكور في المجتمع المدروس.

4-5: تقدير الخطأ المعياري لتقدير متوسط المجتمع المجهول $\bar{y}$:

إن تقدير متوسط المجتمع بواسطة متوسط العينة $\bar{x}$ يختلف من عينة لأخرى. وبما أنه يمكننا أن نسحب من ذلك المجتمع عدداً كبيراً من العينات وبنفس الحجم n , فإن ذلك يعطينا عدداً كبيراً من التقديرات لمتوسط المجتمع $\bar{y}$، وأن هذه التقديرات تختلف عن بعضها البعض, ويكون لها متوسط وتباينٍ وانحراف معياري.

فكيف نقدر الانحراف المعياري للتقدير الذي استخدمناه في تقدير متوسط المجتمع $\bar{y}$؟

ومن المعلوم في الإحصاء أن متوسط الخطأ المرتكب من جراء تقدير متوسط المجتمع بواسطة العينة، يسمى بالخطأ المعياري أو بالانحراف المعياري لمتوسط العينة $\bar{x}$ ويرمز له بـ $\sigma_{\bar{x}}$ ، وهو يقدر حسب حالة السحب من خلال تقدير التباين s^2، بواسطة العلاقة التالية:

$$\tilde{\sigma}_{\bar{x}}^2 = \begin{cases} \dfrac{s^2}{n} & \text{في حالة السحب مع الإعادة} \\ \dfrac{N-n}{N} \cdot \dfrac{s^2}{n} & \text{في حالة السحب بدون إعادة} \end{cases} \qquad (4-13)$$

ملاحظة هامة: إذا كان حجم المجتمع N كبيراً، فإن الكسر $\frac{N-n}{N-1} \approx 1$، وعندها يمكننا اعتبار المقدار $\frac{s^2}{n}$، وفي كلتا حالتي السحب، تقديراً للتباين $\tilde{\sigma}_{\bar{x}}^2$. وبذلك نحصل على أن التقدير $\tilde{\sigma}_{\bar{x}}^2$ عندما يكون حجم المجتمع كبيراً يعطى بالعلاقة:

$$\tilde{\sigma}_{\bar{x}}^2 = \frac{s^2}{n} \qquad (4-14)$$

ومنها نقدر الخطأ المعياري $\sigma_{\bar{x}}$ بالعلاقة التالية:

$$\tilde{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}} \qquad (4-15)$$

مثال(4-3): أن تقدير الخطأ المعياري $\sigma_{\bar{x}}$ لتقدير متوسط العمر $\bar{x}$ المحسوب من معطيات العينة في المثال السابق (4-1), وبفرض أن حجم المجتمع كان كبيراً, يساوي:

$$\tilde{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{2.10}{\sqrt{10}} = 0.66$$

أي أن متوسط العمر في المجتمع $\bar{y}$ يقدر بـ 23 عاماً، وبخطأ معياري قدره 0.66 عاماً.

وهذا يعني أن الخطأ المعياري الذي وقع عند تقدير متوسط العمر $\bar{y}$ يقدر بـ 0.66 عاماً.

4-6: تقدير نسبة خاصة معينة في المجتمع:

لنفترض أننا نريد تقدير نسبة المدخنين في المجتمع المدروس من خلال عينة حجمها n فرداً. وعند دراسة أفراد تلك العينة وجدنا أن عدد المدخنين فيها يساوي m فرداً.

وبذلك نجد أن نسبة المدخنين في العينة، والتي سنرمز لها بالحرف r تساوي:

$$r=\frac{m}{n} \qquad (16-4)$$

وهكذا تكون نسبة غير المدخنين والتي سنرمز لها بـ q تساوي:

$$q=\frac{n-m}{n}=1-r \qquad (17-4)$$

وبناءً على ذلك نقوم بتقدير نسبة المدخنين في المجتمع، والتي سنرمز لها بـ R، بواسطة نسبتهم في العينة r. ونكتب ذلك على الشكل التالي:

$$\tilde{R}=r=\frac{m}{n} \qquad (18-4)$$

ويمكن البرهان على أن الخطأ المعياري للتقدير r، والذي سنرمز بـ $\boldsymbol{\sigma_r}$ يقدر بواسطة العلاقة:

$$\tilde{\sigma}_r=\sqrt{\frac{r\cdot q}{n}} \qquad (19-4)$$

علماً أن: q=1-r، وذلك لأن التباين المصحح للنسبة المدروسة يقدر بـ: $s^2=r\cdot q$

مثال(4-4):

لتقدير نسبة المدخنين بين الموظفين في الجامعة، سحبنا عينة عشوائية منهم بحجم n=160 موظفاً، فوجدنا أن 35 منهم يدخنون. وبذلك نجد أن نسبة المدخنين في العينة تساوي:

$$r=\frac{m}{n}=\frac{35}{160}=0.2188=(21.88\%)$$

وبذلك نجد أن نسبة المدخنين في مجتمع الموظفين تقدر بما يلي:

$$\tilde{R}=r=0.2188$$

وأن الخطأ المعياري في تقدير r يقدر بـ:

$$\tilde{\sigma}_r=\sqrt{\frac{r\cdot q}{n}}=\sqrt{\frac{(0.2188)(1-0.2188)}{160}}$$
$$=\sqrt{0.0010683}=0.0327$$

أي أن نسبة المدخنين في المجتمع المذكور تقدر بـ 21.88% وبخطأ معياري قدره 3.27%.

4-7: تقدير الفرق بين متوسطي مجتمعين طبيعيين:

لنفترض أننا نريد مقارنة متوسطي المجتمع، لمتحول واحد معين X في مجتمعين مختلفين، لذلك نسحب عينة من المجتمع الأول بحجم n_1 وعينة من المجتمع الثاني بحجم n_2, ولنفترض أن متوسطيهما كانا $\bar{x}_1$ و

$\bar{x}_2$ على الترتيب. وإذا رمزنا لمتوسط المجتمع الأول بـ $\bar{y}_1$ ، ولمتوسط المجتمع الأول بـ $\bar{y}_2$، وبناءً على ماتقدم يمكننا أن نقدر كلاً من هذين المتوسطين بواسطة متوسط العينة المقابلة له. أي يكون لدينا:

$$\tilde{\bar{y}}_1 = \bar{x}_1 \qquad (20-4)$$

$$\tilde{\bar{y}}_2 = \bar{x}_2$$

وبذلك يمكننا ببساطة أن نقدر الفرق بين المتوسطين $(\bar{y}_1 - \bar{y}_2)$ بواسطة الفرق بين متوسطي العينتين $(\bar{x}_1 - \bar{x}_2)$، ونكتب ذلك كما يلي (مع الحفاظ على الترتيب):

$$\widetilde{(\bar{y}_1 - \bar{y}_2)} = \bar{x}_1 - \bar{x}_2 \qquad (21-4)$$

ويقدر الخطأ المعياري لهذا التقدير للفرق بواسطة جذر مجموعي التباينين المتعلقين بتقدير كل من $\bar{x}_1, \bar{x}_2$ (لأن الأخطاء لا تطرح بلا تجمع) وهو يساوي:

$$\tilde{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\tilde{\sigma}^2_{\bar{x}_1} + \tilde{\sigma}^2_{\bar{x}2_1}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \qquad (22-4)$$

وبطريقة مشابهة يمكننا تقدير الفرق بين نسبتين في مجتمعين $(R_1\text{-}R_2)$ بواسطة الفرق بين النسبتين في العينتين ، واللتين سنرمز لهما بـ $(r_1\text{-}r_2)$. فيكون لدينا (مع الحفاظ على الترتيب):

$$(\widetilde{R1 - R2}) = (r1 - r2) \qquad (23-4)$$

وإن تقدير الخطأ المعياري لذلك التقدير يحسب من خلال العلاقة التالية:

$$\tilde{\sigma}_{\bar{r}_1 - \bar{r}_2} = \sqrt{\tilde{\sigma}^2_{\bar{r}_1} + \tilde{\sigma}^2_{\bar{r}2_1}} = \sqrt{\frac{r_1 \cdot q_1}{n_1} + \frac{r_1 \cdot q_1}{n_2}} \qquad (24-4)$$

مثال(4-5):

نفترض أننا نريد تقدير الفرق بين متوسطي الدخل الشهري للصيادلة في مدينتين، وتقدير الفرق بين نسبتي المدخنين فيهما، فسحبنا من كل منهما عينتين الأولى بحجم $n_1=30$ ، والثانية بحجم $n_2=20$ ، فكان متوسط الدخل في العينة الأولى $\bar{x}_1 = 20000$ ، ومتوسطه في العينة الثانية $\bar{x}_2 = 15000$ ل.س /شهريا. وكان الانحراف المعياري للعينة الأولى $s_1 = (360)$ والانحراف المعياري للعينة الثانية $s_2 = (150)$, كما كانت نسبة المدخنين فيهما $r_1 = 0.35$ و $r_2 = 0.27$ على الترتيب.

فكيف يمكننا أن نقارن بين هذين المجتمعين ونقدر الفروقات بينهما؟

للإجابة على ذلك نقوم بتقدير الفرق بين متوسطي الدخل من العلاقة (4-21)،ص فنجد أن:

$$\widetilde{(\bar{y}_1 - \bar{y}_2)} = \bar{x}_1 - \bar{x}_2 = 20000 - 15000 = 5000$$

أما الخطأ المعياري لذلك التقدير فيحسب من العلاقة (4-22), وبذلك نجد أنه يساوي:

$$\tilde{\sigma}_{\bar{x}_1-\bar{x}_2}=\sqrt{\tilde{\sigma}^2_{\bar{x}_1}+\tilde{\sigma}^2_{\bar{x}2_1}}=\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

$$=\sqrt{\frac{(360)^2}{30}+\frac{(150)^2}{20}}$$

$$=\sqrt{4320+1125}=73.79$$

أي أن الفرق بين متوسطي الدخل في هاتين المدينتين يقدر بـ 5000 ل.س، وبخطأ معياري قدره 73.79 ليرة.

أما بالنسبة لتقدير الفرق بين نسبتي المدخنين فنستخدم العلاقة التالية:

$$(\widetilde{R1-R2})=(r1-r2)=0.35-0.27=0.08$$

ولتقدير الخطأ المعياري المتعلق بذلك الفرق نطبق العلاقة التالية:

$$\tilde{\sigma}_{\bar{r}_1-\bar{r}_2}=\sqrt{\frac{r_1\cdot q_1}{n_1}+\frac{r_2\cdot q_2}{n_1}}=\sqrt{\frac{(0.35)(1-0.35)}{30}+\frac{(0.27)(1-0.27)}{23}}$$

$$=\sqrt{0.0076+0.0099}=0.1323$$

أي أن الفرق بين نسبتي المدخنين في المدينتين يقدر بـ 0.08 وبخطأ معياري كبير يساوي 0.1323 .

8-4: إنشاء مجالات الثقة لمعالم المجتمع:

بعد أن درسنا عدداً من التقديرات النقطية لبعض مؤشرات المجتمع وتوصلنا إلى علاقات رياضية لحساب كل منها، نشير إلى أن تلك التقديرات غير متحيزة وفعالة ومتماسكة للمؤشر المجهول، إلا أن أياً منها لا يعطينا أية درجة من الثقة فيه. فنحن لانعرف فيما إذا كانت قيمة التقدير النقطي قريبة من القيمة الحقيقية المجهولة أم بعيدة عنها. ومن هنا كان لابد من إيجاد وسيلة تضمن لنا وباحتمال كبير أن تكون تلك القيمة الحقيقية المجهولة واقعة ضمن مجال معين حول القيمة المقدرة.

لذلك سنقوم بإنشاء مجالات الثقة, التي تضمن لنا وقوع القيمة الحقيقية المجهولة للمؤشر المدروس ضمنها باحتمالات كبيرة. وسنميز هنا بين حالتين لحجم العينة (عينة كبيرة وعينة صغيرة):

1 . إذا كان حجم العينة كبيراً $(\mathbf{n \geq 30})$:

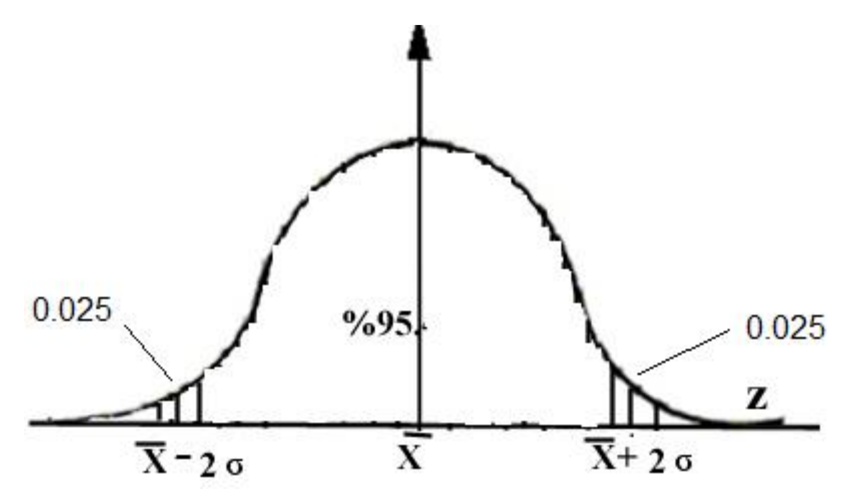
في هذه الحالة يتم إنشاء مجالات الثقة المختلفة لمتوسط المجتمع المجهول $\bar{y}$ اعتماداً على التوزيع الطبيعي المعياري. وبذلك نجد أن مجال الثقة للمتوسط $\bar{y}$ ، يعطى بالعلاقة التالية:

$$\mathbf{P}[\bar{x}-Z\,\widetilde{\sigma_{\bar{x}}}\leq\bar{y}\leq\bar{x}+Z\,\widetilde{\sigma_{\bar{x}}}]$$

وبما أن الخطأ المعياري لمتوسط العينة $\widetilde{\sigma_{\bar{x}}}$ يقدر بالعلاقة : $\widetilde{\sigma}_{\bar{x}}=\frac{s}{\sqrt{n}}$ ، فإننا نجد أن مجال الثقة الثاني (عندما نضع (Z=1,96=2 لمتوسط المجتمع $\bar{y}$ يعطى بالعلاقة :

$$\mathbf{P}[\bar{x}-2\,.s/\sqrt{n}\leq\bar{y}\leq\bar{x}+2\,\,.s/\sqrt{n}]=\mathbf{0.9545}\qquad(25-4)$$

وهو مجال يحوي القيمة الحقيقية المجهولة $\bar{y}$ باحتمال 0.9545، أي بمستوى دلالة α يساوي: $\alpha=$ 0.05 للطرفين(0,025 لكل طرف) . كما هو مبين في الشكل (4-1).



الشكل (4−1): مجال الثقة الثاني

أما مجال الثقة الثالث فيعطى بالعلاقة التالية :

$$P[\bar{x} - (3)\ \widetilde{\sigma}_{\bar{x}} \leq \bar{y} \leq \bar{x} + (3)\ \widetilde{\sigma}_{\bar{x}}] = 0.9972 \qquad (26-4)$$

وهو مجال يحوي القيمة الحقيقية المجهولة $\bar{y}$ باحتمال قدره 0.9972 ، أي بمستوى دلالة α=0.0028 للطرفين(0,0014 لكل طرف).

أما بالنسبة لنسبة خاصة في المجتمع فيمكن إنشاء مجال الثقة الثاني للنسبة R من العلاقات التالية:

$$?\mathbf{P[r - 2\ \widetilde{\sigma}_r \leq R \leq r + 2\ \widetilde{\sigma}_r] = P\left[r - 2\sqrt{\frac{r\cdot q}{n}} \leq R \leq r + 2\sqrt{\frac{r\cdot q}{n}}\right] = 0.95 \quad (27-4)}$$

وهو المجال الذي يحوي القيمة الحقيقية R باحتمال يساوي 0.95 ، أي بمستوى دلالة يساوي α=0.05 للطرفين.

مثال(4−6): أوجد مجال الثقة الثاني لمتوسط الإنفاق الشهري على المكالمات بالهاتف الجوال لطلاب الجامعة، علماً بأنه سحبنا عينة عشوائية منهم بحجم n=50، وكان متوسط الانفاق $\bar{x}$ =1220، وانحرافه المعياري المصحح s=755 . الجل: من المعطيات نجد أن:

$$\tilde{\bar{y}} = \bar{x} = 1220$$
$$\widetilde{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{755}{\sqrt{50}} = 106.7$$
$$P[\bar{x} - 2\widetilde{\sigma}_{\bar{x}} \leq \bar{y} \leq \bar{x} + 2\widetilde{\sigma}_{\bar{x}}] =$$
$$= P[1220 - 2(106.7) \leq \bar{y} \leq 1220 + 2(106.7)]$$
$$= P[1220 - 213.5 \leq \bar{y} \leq 1220 + 213.5]$$
$$= P[1006.5 \leq \bar{y} \leq 1433.54] = 0.95$$

أي أن متوسط الإنفاق الحقيقي الشهري $\bar{y}$ للطلاب في الجامعة على الاتصالات بالهاتف الجوال ، يقع في المجال [1006.5، 1433.54] وباحتمال قدره 0.95 .

وكذلك يمكننا أن نجد مجال الثقة الثاني لنسبة المدخنين الواردة في المثال (4−4) حيث نجد أن ذلك المجال يساوي:؟

$$P[r - 2\widetilde{\sigma}_r \leq R \leq r + 2\widetilde{\sigma}_r] =$$

$= P[0.2188 - 2(0.0327) \leq R \leq 0.2188 + 2(0.0327)]$
$= P[0.1534 \leq R \leq 0.2842] = 0.95$

وهو المجال الذي يحوي النسبة الحقيقية المجهولة R باحتمال قدره 0.95 .

2 . إذا كان حجم العينة صغيراً ($\mathbf{n} < 30$):

في هذه الحالة يتم إنشاء مجالات الثقة الثاني لمتوسط المجتمع المجهول $\bar{\mathbf{y}}$ اعتماداً على توزيع ستودينت (توزيع t)، وبذلك نجد أن مجال الثقة لمتوسط المجتمع $\bar{\mathbf{y}}$ ، يعطى في هذه الحالة بواسطة العلاقة التالية:

$$\mathbf{P}[\bar{\mathbf{x}} - \mathbf{t} \cdot \tilde{\boldsymbol{\sigma}}_{\bar{\mathbf{x}}} \leq \bar{\mathbf{y}} \leq \bar{\mathbf{x}} + \mathbf{t} \cdot \tilde{\boldsymbol{\sigma}}_{\bar{\mathbf{x}}}] = \mathbf{P}\left[\bar{\mathbf{x}} - \mathbf{t} \cdot \frac{\mathbf{s}}{\sqrt{\mathbf{n}}} \leq \bar{\mathbf{y}} \leq \bar{\mathbf{x}} + \mathbf{t} \cdot \frac{\mathbf{s}}{\sqrt{\mathbf{n}}}\right] \qquad (\mathbf{28-4})$$

حيث أن t هي القيمة الجدولية لمتحول توزيع ستودينت المقابلة لمستوى دلالة 0.05 أو 0.10 ولعدد درجات الحرية قدره (n-1)، ويتم البحث عنها في جداول التوزيع t .

وبطريقة مشابهة نجد أن مجال الثقة للنسبة الحقيقية R في المجتمع يعطى بالعلاقة الآتية:

$$\mathbf{P}[\mathbf{r} - \mathbf{t} \cdot \tilde{\boldsymbol{\sigma}}_{r} \leq \bar{\mathbf{y}} \leq \mathbf{r} + \mathbf{t} \cdot \tilde{\boldsymbol{\sigma}}_{r}] = \mathbf{P}\left[\mathbf{r} - \mathbf{t} \cdot \frac{\sqrt{\mathbf{r} \cdot \mathbf{q}}}{\sqrt{\mathbf{n}}} \leq \mathbf{R} \leq \mathbf{r} + \mathbf{t} \cdot \frac{\sqrt{\mathbf{r} \cdot \mathbf{q}}}{\sqrt{\mathbf{n}}}\right] \qquad (\mathbf{29-4})$$

مثال(4−7): لنأخذ المعطيات الواردة في المثال (4−2) حول متوسط العمر المأخوذ من عينة حجمها n=10 ومتوسطها $\bar{\mathbf{x}} = \mathbf{23}$ وخطأها المعياري $\tilde{\boldsymbol{\sigma}}_{\bar{\mathbf{x}}} = \mathbf{0.66}$. والمطلوب إنشاء مجال الثقة الثاني لمتوسط العمر الحقيقي $\bar{\mathbf{y}}$ بمستوى دلالة $\boldsymbol{\alpha} = \mathbf{0.10}$:

بما أن حجم العينة صغير نحسب مجال الثقة اعتماداً على العلاقة (4−27) فنجد أن:

$\boldsymbol{P}[\bar{\mathbf{x}} - \mathbf{t} \cdot \tilde{\boldsymbol{\sigma}}_{\bar{\mathbf{x}}} \leq \bar{\mathbf{y}} \leq \bar{\mathbf{x}} + \mathbf{t} \cdot \tilde{\boldsymbol{\sigma}}_{\bar{\mathbf{x}}}] =$
$\mathbf{P}[\mathbf{23} - \mathbf{t} \cdot (\mathbf{0.66}) \leq \mathbf{y} \leq \mathbf{23} + \mathbf{t} \cdot (\mathbf{0.66})]$

ومن جدول توزيع ستيودينت الملحق نجد أن قيمة t المقابلة لمستوى دلالة 0.10 للطرفين (0.05 لكل طرف) ولدرجة حرية $n\text{-}1=9$ تساوي $t=1.833$. وبذلك نجد أن المجال المطلوب هو:

$$\mathbf{P}[\mathbf{23} - \mathbf{1.833} \cdot (\mathbf{0.66}) \leq \mathbf{y} \leq \mathbf{23} + \mathbf{1.833} \cdot (\mathbf{0.66})] = \mathbf{P}[\mathbf{21.79} \leq \mathbf{y} \leq \mathbf{24.21}] = \mathbf{0.90}$$

9−4 : حساب حجم العينة n :

إن حجم العينة اللازم سحبه في دراسة معينة يتوقف على مقدار تجانس المجتمع بالنسبة للخاصة المدروسة، وعلى مقدار الدقة المطلوبة d ، وعلى معامل الثقة المفروض z ، وعلى مقادير أخرى كالتكلفة وطريقة السحب. **وهو بصورة عامة يجب أن يحقق الشروط التالية:**

1 . أن يتناسب طرداً مع احتمال الثقة المطلوب، أو مع مربع معامل احتمال **z** الذي يؤخذ من جدول التوزيع الطبيعي.

2 . أن يتناسب طرداً مع تباين العينة s^2, لأن ذلك التباين يعبّر عن تشتت القيم في العينة والمجتمع، فكلما كان التشتت كبيراً لزم أن يكون حجم العينة كبيراً.

3 . أن يتناسب عكساً مع الدقة المرغوبة $\boldsymbol{d}$ أو مربعها $\boldsymbol{d^2}$. علماً بأن الدقة $\boldsymbol{d}$ هي الحد الأعلى للخطأ المسموح به. ويتم تحديده مسبقاً من قبل المسؤولين عن البحث أو الدراسة.

4. إن حساب حجم العينة يتعلق بنوع المؤشر المراد تقديره : كالمتوسط والنسبة وغيرهما.

4-9-1: تقدير حجم العينة لتقدير المتوسط:

عندما نريد تقدير متوسط المجتمع $\bar{y}$ بدقة قدرها d وباحتمال ثقة قدره β، فإننا نجد اعتماداً على ما سبق أن نصف طول مجال الثقة يجب أن يكون أصغر من تلك الدقة, أي أن:

$$Z_{1-\frac{\alpha}{2}} \tilde{\sigma}_{\bar{x}} \leq d$$

وبتربيع الطرفين نجد أن:

$$Z^2_{1-\frac{\alpha}{2}} \tilde{\sigma}^2_{\bar{x}} \leq d^2 \qquad (30-4)$$

وبما أن قيمة $\tilde{\sigma}^2_{\bar{X}}$ ترتبط بحالة السحب فإننا سنميز بين حالتي السحب فنجد أنه في:

1 . حالة السحب مع الإعادة: حيث يكون لدينا:

$$\tilde{\sigma}^2_{\bar{X}} = \frac{s^2}{n}$$

وبالتعويض في المتراجحة السابقة نحصل على:

$$Z^2_{1-\frac{\alpha}{2}} \cdot \tilde{\sigma} \cdot \frac{s^2}{n} \leq d^2$$

ومنها نجد أن حجم العينة ***n*** يجب أن يحقق في هذه الحالة المتراجحة التالية:

$$n \geq \frac{Z^2_{1-\frac{\alpha}{2}} \cdot s^2}{d^2} \qquad (31-4)$$

ومنها يمكننا حساب الحد الأدنى لحجم العينة من العلاقة:

$$n_0 = \frac{Z^2 \cdot s^2}{d^2} \qquad (32-4)$$

2 . حالة السحب بدون إعادة: في هذه الحالة يكون لدينا:

$$\tilde{\sigma}^2_{\bar{X}} = \frac{N-n}{N} \cdot \frac{s^2}{n}$$

وبالتعويض في (4-30) نحصل على أن:

$$Z^2 \cdot \frac{N-n}{N} \cdot \frac{s^2}{n} \leq d^2$$

وبإصلاح هذه المتراجحة نجد أن حجم العينة ***n*** في حالة السحب بدون إعادة يجب أن يحقق العلاقة التالية:

$$n \geq \frac{N \cdot Z^2 \cdot s^2}{Nd^2 + Z^2 s^2} \qquad (33-4)$$

ومنها يمكننا حساب الحد الأدنى لحجم العينة في حالة السحب بدون إعادة من العلاقة:

$$n_0 = \frac{N \cdot Z^2 \cdot s^2}{Nd^2 + Z^2 \cdot s^2} \qquad (34-4)$$

وبذلك نكون قد توصلنا إلى علاقتين تعطياننا حجم العينة ***n*** في كلتا حالتي السحب. إلا أن s^2 الداخل فيهما يحسب من تباين أي عينة اختبارية تسحب على سبيل التجربة.

ملاحظة: يمكن استخدام العلاقتين (4–32) و(4–34) عند تقدير حجم العينة لإجمالي المجتمع ***Y*** ، مع الانتباه إلى ضرب دقة الإجمالي بـ ***N***.

مثال(4–8): أوجد حجم العينة اللازم لتقدير متوسط دخل الأسرة بدقة $d = 200$ ل.س. وباحتمال ثقة قدره 0,95 ، علماً بأن عينة اختبارية أعطتنا أن $s^2 = 300000$ وأن السحب سيجري مع الإعادة.

الحــل: لحساب حجم العينة ***n*** نطبق من العلاقة (4–32) فنجد أن قيمة ***Z*** المقابلة للاحتمال $1-\frac{\alpha}{2}=0{,}975$ تساوي ***1,96*** ، وهكذا نجد أن:

$$n \geq \frac{Z^2 \cdot s^2}{d^2} = \frac{(1{,}96)^2(300000)}{(200)^2} = 28{,}8 \approx 29$$

أي أنه يكفي أن تكون $n = 29$ لنحقق الدقة المطلوبة وباحتمال الثقة المطلوب.

مثال 14: أوجد حجم العينة اللازم لتقدير متوسط استهلاك الأسرة من المازوت في الشهر وبدقة $d = 20$ لتراً، وباحتمال قدره 0,90 ، علماً بأن عينة اختبارية أعطتنا أن: $s^2 = 16000$ ، وأن السحب سيجري بدون إعادة، وأن حجم المجتمع المدروس $N = 1000$ أسرة.

الحل: بما أن السحب يجرى بدون إعادة فإننا نطبق العلاقة (4–34)، ومن جدول التوزيع الطبيعي المعياري نجد أن قيمة ***Z*** المقابلة للاحتمال $1-\frac{\alpha}{2}=0{,}95$ تساوي $Z = 1{,}65$ ، وبذلك نجد أن:

$$n \geq \frac{N \cdot Z^2 \cdot s^2}{N \cdot d^2 + Z^2 s^2} = \frac{1000(1{,}65)^2(16000)}{100(20)^2 + (1{,}65)^2(16000)} = 98{,}2 \approx 99`$$

أي أنه يكفي أن يكون $n = 99$ أسرة لنحقق الدقة المطلوبة وبالاحتمال المطلوب.

ملاحظة: إن حجم العينة الذي نحصل عليه هو الحد الأدنى لحجم العينة *n*، ويمكن أن يكون حجمها أكبر من ذلك. ونشير هنا إلى أن حجم العينة لابد وأن يكون عدداً صحيحاً. لذلك يجب تقريبه إلى الأعلى دائماً.

4–9–2: تقدير حجم العينة لتقدير النسبة:

عندما نريد تقدير نسبة خاصة ما في مجتمع ***R*** وبدقة δ وباحتمال ثقة قدره $(1-\alpha)$، فإننا نجد اعتماداً على العلاقة (4–29)، إنه إذا كانت δ هي الدقة النسبية المطلوبة فإن:

$$Z_{1-\alpha/2}\tilde{\sigma}_r \leq \delta$$

$$Z^2_{1-\alpha/2}\tilde{\sigma}^2_r \leq \delta^2$$

وبما أن $\tilde{\sigma}^2_r$ ترتبط بحالة السحب فإننا نجد أنه:

1 . في حالة السحب مع الإعادة: يكون لدينا:

$$Z^2_{1-\alpha/2} \cdot \frac{rq}{n} \leq \delta^2$$

ومنها نحصل على أن:

$$n \geq \frac{Z^2_{1-\alpha/2} \cdot rq}{\delta^2} \qquad (35-4)$$

2 . في حالة السحب بدون إعادة: يكون لدينا:

$$Z^2_{1-\alpha/2} \cdot \frac{N-n}{N} \cdot \frac{S^2}{n} \leq \delta^2 \qquad (36-4)$$

ومنها نحصل على أن:

$$n \geq \frac{NZ^2_{1-\alpha/2} rq}{N\delta^2 + Z^2_{1-\alpha/2} rq} \qquad (37-4)$$

علماً بأن r و $(q=1-r)$ تحسب بشكل مبدئي من أية عينة اختبارية سريعة.

مثال(4-9): أوجد حجم العينة اللازم سحبها لتقدير نسبة المدخنين بدقة $\delta = 0{,}10$، وباحتمال قدره 0,95 ، علماً بأن السحب سيجري مع الإعادة، وأن عينة اختبارية سريعة أعطتنا أن نسبتهم $r = 0{,}4$.

الحل: من العلاقة (4-35) نجد أن:

$$n \geq \frac{(1{,}96)^2 (0{,}4)(0{,}6)}{(0{,}01)^2} = 256{,}1 \approx 257$$

أي أنه يكفي أن يكون $n = 257$ لنحقق الدقة المطلوبة وبالاحتمال المطلوب.

مثال(4-10): أوجد حجم العينة اللازم لتقدير متوسط استهلاك الأسرة من المياه في المجتمع شهرياً، ثم حجمها لتقدير نسبة المدخنين فيه، إذا علمت أننا نريد أن يكون ذلك التقدير باحتمال ثقة قدرة 0.95 (أي بمستوى دلالة أقل من 0.05)، وأن الدقة المطلوبة في تقدير متوسط الاستهلاك تساوي **d=5** غالون شهرياً وفي تقدير نسبة المدخنين **d=0.10** , وأن العينة الاختبارية أعطتنا أن $s^2 = 90$, وأن نسبة المدخنين فيها كانت تساوي **r = 0.60**.

الحل: من الشروط السابقة نجد أن قيمة **z** المقابلة لاحتمال الثقة المذكور هي: **z=2** تقريباً.

وأن حجم العينة اللازم لتقدير المتوسط يُحسب من العلاقة الآتية(لأن المجتمع كبير):

$$n_0 = \frac{z^2 s^2}{d^2} = \frac{(2)^2 \cdot 90}{(5)^2} = 14.4 \approx 15$$

أي أن الحد الأدنى لحجم العينة الذي يحقق الشروط السابقة لتقدير متوسط الاستهلاك هو15 أسرة .

أما بالنسبة لحجم العينة اللازم لتقدير نسبة المدخنين R فنحسبه من العلاقة:

$$n_0 = \frac{z^2 \cdot r \cdot q}{d^2} = \frac{(2)^2 \cdot (0.60) \cdot (0.40)}{(0.10)^2} = 96$$

وهو الحد الأدنى لحجم العينة اللازم لتحقيق الشروط السابقة في تقدير النسبة **R**.

تمرينات

1 . لنفترض أن مجتمعاً مؤلفاً من العناصر المعلومة التالية:

$$Yi:12,14,16,18,15,17,16,19,15,17$$

. اسحب بشكل عشوائي وبدون إعادة عينة بحجم $n=3$.

. أوجد تقديراً لمتوسط المجتمع وقارنه مع المتوسط الحقيقي $\bar{y}$.

. أوجد تقديراً لتباين ذلك المجتمع وقارنه مع التباين الحقيقي.

. أوجد مجال ثقة لمتوسط المجتمع وباحتمال قدره $0,95$.

. حاول أن تسحب عينة أخرى وبنفس الحجم ثم أوجد مختلف التقديرات السابقة.

2 . سحبنا عينة بدون إعادة وبحجم $n=3$ من مجتمع مؤلف من 50 أسرة فوجدنا أن دخولها الشهرية تساوي (ألف ل.س).

$$X_i:12,10,14,18,17,11,19,20,13$$

والمطلوب:

. إيجاد تقدير لمتوسط الدخل في المجتمع $\bar{y}$.

. إيجاد تقدير لتباين الدخل فيه σ^2.

. إيجاد تقدير للتباين $\sigma_{\bar{x}}^2$.

. إيجاد تقدير نسبة الدخول التي هي أقل من 15 وتقدير تباينها.

. إيجاد مجال الثقة المقابل للاحتمال $0,95$ لكل من المتوسط والنسبة.

3 . لنفترض أن مجتمعاً مؤلفاً من 500 عنصراً. ونريد أن نسحب منه مع الإعادة عينة عشوائية لتقدير متوسطه بدقة $d=5$، وباحتمال قدره $0,95$. فماهو حجم العينة اللازم تحديده لتأمين الدقة المطلوبة وبالاحتمال المفروض؟، وذلك إذا علمت أن تباينه يقدر ب: $s^2=250$.

4. احسب حجم العينة اللازم لتحقيق دقة d=0.01، وباحتمال 0.95 في حالة السحب مع الإعادة عند تقدير نسبة المدخنين علماً بأن التقدير الأولي لنسبة المدخنين كان r=0.25.

5 . سحبنا مع الإعادة من بين الطالبات عينة بحجم $n=15$ طالبة، ومن بين الطلاب عينة بحجم $n=20$ طالباً، ودرسنا متوسط نفقاتهم على مستلزماتهم الجامعية في الشهر، فوجدنا أن متوسط نفقات الطالبة في العينة $\bar{x}_1=1000$ ل.س ، وأن متوسط نفقات الطالب في العينة $\bar{x}_2=1500$ ل.س، وأن تباين النفقات في عينة الطالبات كان يساوي $s_1^2=600$ ل.س ، وأن تباين النفقات في عينة

الطلاب كان يساوي $s_1^2 = 500$ ؟ل.س ، وأن نسبة ما تنفقه الطالبات في العينة على الكتب كانت تساوي $r_1 = 0{,}10$ ، بينما كانت نسبة ما ينفقه الطلاب في العينة على الكتب تساوي $r_1 = 0{,}20$ ، والمطلوب:

. تقدير الفرق بين متوسطي النفقات للطالبات والطلاب ثم إيجاد مجال الثقة لذلك باحتمال قدره $0{,}95$

. تقدير الفرق بين نسبتي النفقات على الكتب ثم إيجاد مجال الثقة لذلك باحتمال قدره $0{,}95$.

الفصل الخامس
تصميم وتحليل الاستبيان

5-1: تمهيد :

إن الاستبيان هو أداة لجمع البيانات عن أحوال عناصر موضوع الدراسة من خلال عينة مسحوبة عشوائياً من مجتمع البحث، ويكون الاستبيان مؤلفاً من عدة محاور، كل منها يتضمن أسئلة أو عبارات محددة عن جوانب الموضوع. ويقوم أفراد العينة بالإجابة عليها، بالطريقة المباشرة (المقابلة) أو عبر البريد العادي أو الالكتروني. وهناك شروط معينة لتصميم الاستبيان ولعدد الخيارات الممكنة للإجابة على الأسئلة الواردة فيه. وعادة ما يتم تقديم الاستبيان بتعليمات حول كيفية التعامل مع الأسئلة والإجابة عليها.

ويبدأ بطرح بعض الأسئلة عن أحوال الشخص المبحوث (جنسه + عمره+ تعليمه+ عمله ...الخ).

أما جسم الاستبيان فيتألف من عدة محاور تعبر عن المتحولات المعتمدة في البحث (لكل متحول محور)، ويتضمن كل محور عدداً محدداً من الأسئلة المباشرة أو العبارات الواضحة، التي تعبر عن ذلك المحور أو تشكل جزءاً منه، ويجب أن تكون الأسئلة أو العبارات ضمن المحاور قصيرة وواضحة وذات اتجاه واحد، وتتناسب مع مستوى المبحوثين، ولا تتضمن عبارات محرجة أو جارحة أو مسيئة أو سخيفة، ولا توحي للمبحوث باختيار إجابة معينة.

أما خيارات الأجوبة فيمكن أن تكون مغلقة أو مفتوحة .

وأهم الخيارات المغلقة هي خيارات (ليكرت) التالية:

1- الخيار الثنائي: ويكون أمام المبحوث خياران فقط للجواب على السؤال، مثل:

نعم	لا
1	0

2- الخيار الثلاثي: ويكون للسؤال ثلاثة خيارات للجواب، مثل:

موافق	محايد	معارض
3	2	1

3- الخيار الخماسي: ويكون للسؤال خمسة خيارات للجواب، مثل:

5	4	3	2	1

4- الخيار السباعي: ويكون للسؤال سبعة خيارات للجواب، مثل:

7	6	5	4	3	2	1

ويمكن أن تكون هذه الخيارات لغوية على الشكل التالي:

موافق جداً	**موافق**	**محايد**	**معارض**	**معارض جداً**
5	**4**	**3**	**2**	**1**

ثم يتم استبدالها بالأرقام المقابلة لها لتحليلها واستخلاص النتائج الممكنة.

وهنا لابد أن نشير إلى أن شكل السؤال أو العبارات يجب أن يتوافق مع شكل الخيارات وبالعكس.

فإذا كان شكل السؤال مباشراً وكمياً فإن الخيارات يجب أن تكون رقمية مثل:

ماهي درجة التزامك بالرياضة ؟	5	4	3	2	1

أما إذا كان السؤال على شكل عبارة استفهامية فإن شكل الخيارات يكون لغوياً مثل:

انت شخص ملتزم بالرياضة؟	موافق جداً	موافق	محايد	معارض	معارض جداً

ثم يتم تحويل هذه الإجابة إلى أرقام مرتبة حسب ما يراه الباحث مناسباً.

كما نشير إلى أن الخيارات الرقمية أفضل وأدق من الخيارات اللغوية، لأن المسافات بين الخيارات الرقمية محددة ومتساوية (وتساوي الواحد)، أما المسافات بين الخيارات اللغوية فهي غير متساوية وغير محددة، فالمسافة بين (موافق جداً) و(موافق) غير معروفة ولا تساوي المسافة بين (موافق) و(محايد). عدا عن أن اتجاه ترتيبها اللغوي يجعلها عرضة للتحيز أثناء الإجابة.

ويجب أن يكون عدد الخيارات في الاستبيان الواحد موحداً. (ثلاثية أو خماسية أو سباعية لجميع الأسئلة). ولا يجوز اعتماد الاستبيان قبل عرضه على عدد من المختصين لتحكيمه وتصويبه، ثم القيام بتجربته وتمريره على عينة استطلاعية (لا تقل عن 30 فرداً) للتأكد من حسن صياغة الأسئلة ومن حسن فهمها ومن صواب الإجابة عليها، ويتم ذلك بحساب معامل الثبات (ألفا كرونباخ) من بيانات العينة الاستطلاعية فإذا كانت قيمته أكبر من (0.70) يمكننا اعتماد الاستبيان وتمريره على أفراد العينة الكلية ذات الحجم المحدد بـ n فرداً . وعند تحليل الثبات لا يجوز دمج الأسئلة ذات الخيارات المختلفة، بل يتم تحليل كل نوع على حده.

مثال (5-1): لنفترض أننا نريد دراسة أثر الهوايات المختلفة على صحة ونفسية كبار السن (أكبر من 70 عاماً) فصممنا استبياناً خاصاً مؤلفاً من (6) محاور (أو أسئلة) عن ممارسة الهوايات الممكنة (يومياً) لهؤلاء الأشخاص وعن حالتهم الصحية، النفسية، وكان على الشكل التالي:

جدول (5-1): الأسئلة والخيارات:

	نص السؤال أو العبارة (جميعها باتجاه واحد)	الخيارات أو الدرجات الممكنة للجواب				
X_1	ماهي درجة ممارستك للرياضة اليومية ؟	5	4 ✓	3	2	1
X_2	ماهي درجة تذوقك واستماعك للموسيقى ؟	5	4	3 ✓	2	1
X_3	ماهي درجة اهتمامك بالأخبار والسياسة ؟	5 ✓	4	3	2	1
X_4	ماهي درجة مواظبتك على المطالعة ؟	5	4 ✓	3	2	1
X_5	ماهي درجة تعاملك مع شبكات التواصل الاجتماعي ؟	5	4	3 ✓	2	1
X_6	ماهي درجة تقييمك لحالتك الصحية والنفسية ؟	5	4 ✓	3	2	1

وهنا نلاحظ أن كل سؤال من هذه الأسئلة يمكن أن يشكل محوراً خاصاً. ويمكننا أن نضع ضمنه عدة أسئلة أو عبارات حسب هدف البحث . وتتم الإجابة على هذه الأسئلة بسرعة وذلك بوضع إشارة معينة (مثل √ أو ×) على الدرجة المناسبة كما في الجدول السابق.

وأخيراً لنفترض إننا مررنا هذا الاستبيان السابق على عينة صغيرة مؤلفة من (10) أفراد (كمثال) ثم قمنا بوضع أجوبتهم حسب الأفراد (في الأسطر) وحسب السؤال (في الأعمدة) فكانت كما في الجدول التالي:

جدول (5–2): بيانات العينة الاستطلاعية (10 أفراد) :

الأسئلة / الأفراد i	X_1 / Y_1	X_2	X_3	X_4	X_5	X_6	T_{1i}	T_{2i}	T_i	المحك $\bar{x}_i$
1	3 /3	2	3	2	2	2	8	6	14	2.33
2	4 /3	3	2	3	4	3	9	10	19	3.17
3	2 /3	2	4	2	3	2	8	7	15	2.50
4	3 /3	4	3	4	4	3	10	11	21	3.50
5	4 /3	3	4	3	3	4	11	10	21	3.50
6	3 /3	2	2	5	4	2	7	11	18	3.00
7	4 /4	3	3	3	2	3	10	8	18	3.00
8	3 /3	4	4	4	4	4	11	12	23	3.83
9	2 /3	3	2	2	2	2	7	6	13	2.17
10	3 /2	3	4	3	2	4	10	9	19	3.17
لدراسة الصدق										
متوسط السؤال $\bar{x}_j$	3.1 /3	2.9	3.1	3.1	3	2.9	9.1	9	18.1	3.017
الانحراف المعياري	0.7379 / 0.4714	0.7379	0.8756	0.9944	0.9429	0.8756	1.524	2.160	3.247	0.541
	$y_i = 0.3194$						$r = 0.540$			

المصدر: افتراضي (الأرقام التي في زوايا خلايا العمود الثاني هي إجابات التجربة الثانية).

وهنا نطرح السؤالين التاليين:

هل هذه الإجابات ثابتة أم إنها تختلف من تجربة لأخرى أو من عينة لأخرى ؟

هل هذه الإجابات صادقة وتقترب من القيم الحقيقية أو المتوقعة لها ؟

للإجابة على هذين السؤالين نحتاج إلى استخدام أساليب مناسبة لقياس كل من الثبات والصدق. وهو ما سنعرضه فيما يلي:

5–2: أساليب قياس الثبات(Reliability) ويسمى أحياناً بـ (الاتساق الداخلي):

يعرف الثبات: بأنه استقرار الإجابات حول قيم معينة وعدم اختلافها كثيراً من تجربة لأخرى أو من عينة لأخرى . ويقصد بالاتساق الداخلي درجة انسجام الإجابات ضمن كل سؤال أو ضمن كل محور أو ضمن الاستبيان ككل . وهناك عدة أساليب لقياس هذا الثبات، أهمها ما يلي:

5-2-1: أسلوب إعادة التجربة (Parallel التوازي): وبحسب هذا الأسلوب يقوم الباحث بإعادة التجربة وتوزيع الاستبيان على نفس أفراد العينة، مع ضمان تحقيق نفس الشروط والظروف السابقة (بدون إعلامهم بهدف الإعادة).

ثم يقوم بتسجيل الإجابات الجديدة مقابل الإجابات السابقة لكل فرد على كل سؤال. ولقياس ثبات النتائج يقوم بمقارنة الاجابات على كل سؤال في التجربتين، ثم يقوم بحساب معامل الارتباط (البيرسوني) بينهما من العلاقة المعروفة التالية:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{(n-1) * \sigma_x * \sigma_y} \qquad (1-5)$$

حيث أن x_i هي قيمة الإجابة السابقة للفرد i على ذلك السؤال، وأن $\bar{x}$ متوسطها وأن σ_x هو انحرافها المعياري، وأن y_i هي قيمة الإجابة اللاحقة للفرد i على ذلك السؤال، وأن $\bar{y}$ هو متوسطها، و σ_y هو انحرافها المعياري . فإذا كانت قيمة هذا المعامل أكبر من (0.70) نعتبر أن ثبات هذه الإجابات مقبولاً أو جيداً ...الخ.

وكتطبيق على ذلك لنأخذ السؤال الأول فقط، ولنفترض أن إجابات الأفراد اللاحقة عليه هي الأرقام المسجلة في زوايا خلايا العمود X_1، والتي رمزنا لها بـ Y_1 , ثم نقوم بحساب معامل الارتباط بين نتائج هاتين التجربتين من (2−1) فنجد أن قيمته تساوي:

$$r_1 = 0.3194 \qquad \text{(للسؤال الأول فقط)}$$

وهي قيمة صغيرة تدل على درجة ثبات ضعيفة للإجابات على السؤال الأول.

وهكذا نفعل مع بقية الأسئلة ونسجل الإجابات اللاحقة مقابل السابقة ونحسب معامل الارتباط لكل سؤال على حدة ونستخلص درجة الثبات لكل منها.

ولكن إذا كان حجم العينة كبيراً (وزوجياً) وكان تسلسل الأفراد في القائمة عشوائياً فيمكننا اتباع أسلوب آخر هو أسلوب التوازي النصفي ، والذي يتلخص بتجزئة إجابات كل سؤال إلى قسمين متساويين، ثم وضع هذين القسمين مقابل بعضهما في عمودين جديدين وحساب معامل الارتباط بينهما، فنحصل على تقدير درجة الثبات لإجابات ذلك السؤال، فمثلاً لو أخذنا إجابات السؤال الأول وقسمناها إلى قسمين (بفرض أن التسلسل عشوائي) كما يلي:

القسم الأول: يتألف من إجابات الأفراد الخمسة الأولى، والقسم الثاني: يتألف من إجابات الخمسة الأخرى ووضعناهما مقابل بعضهما ثم حسبنا معامل الارتباط بينهما لوجدنا أن: $r_1 = 0.4226$ ، وهو معامل صغير أيضاً ويدل على درجة ثبات ضعيفة لإجابات السؤال الأول.

ملاحظة: في الحقيقة أن معامل الارتباط لا يعبر بشكل جيد عن درجة ثبات الإجابات ، لأنه إذا طرحنا (واحد) من الإجابات السابقة فإنها ستصبح غير ثابتة، وإن قيمة معامل الارتباط بين الإجابات السابقة واللاحقة ستكون مساوية للواحد $r = 1$. رغم أن الإجابات أصبحت متحيزة، ولهذا فإننا سنحاول تطبيق

أساليب أخرى لقياس درجة ثبات الإجابات، ويفضل في هذه الحالة (حالة إعادة التجربة) تطبيق معامل التوافق X^2 بين نتائج التجربتين. أو استخدام اختبار الأزواج المتقابلة على الإجابات السابقة واللاحقة وحساب قيمة مؤشر ستودينت t من العلاقة:

$$t = \frac{\bar{d} - \bar{d}_0}{S_d/\sqrt{n}} \qquad (2-5)$$

حيث أن: $\bar{d}$ هو متوسط الفروقات بين نتائج التجربتين $d_i = (x_i - y_i)$.

وأن S_d هو الانحراف المعياري للفروقات d_i، وn حجم العينة.

أما $\bar{d}_0$ فهي قيمة متوسط الفروقات في المجتمع. ونأخذه من فرضية العدمH_0 (انظر العلاقة (6−54)).

5−2−2: أسلوب تجزئة الأسئلة بالمناصفة (Split- Half) أو أسلوب (سبيرمان− براون):

ويعتمد هذا الأسلوب على تجزئة الأسئلة (وليس الإجابات) إلى جزأين متساويين (بفرض أن عدد الأسئلة K زوجي) حسب تسلسلها أو حسب أي معيار آخر (فردي، زوجي) وتشكيل مجموعتين متقابلتين من الأسئلة. ثم نقوم بحساب متوسطات (أو مجاميع) إجابات الأفراد في أسئلة كل مجموعة، ونضعهما مقابل بعضهما، ثم نحسب معامل الارتباط بينها، وليكن مساوياً لـ r . ثم نقوم بحساب معامل الثبات الذي اقترحه (براون) لزيادة قيمة معامل الثبات وسُمي بمعامل (سبيرمان− براون) وهو يعرف بالعلاقة :

$$BC = \frac{2r}{1+r} \qquad (3-5)$$

ملاحظة: إذا كان عدد الأسئلة K فردياً يتم تجزئتها إلى $\frac{k+1}{2}$ سؤالاً ثم إلى $\frac{k-1}{2}$ سؤالاً (بفارق سؤال واحد)

ولقد قمنا بتجزئة أسئلة الاستبيان الوارد في الجدول (5−2) إلى جزأين متساويين كمايلي:

الجزء الأول: ويتألف من الأسئلة: الأول والثاني والثالث ، ورمزنا لمجموع إجابات الفرد i فيه (بدلاً من متوسطها) بالرمز T_{1i} .

الجزء الثاني: ويتألف من الأسئلة: الرابع والخامس والسادس، ورمزنا لمجموع إجابات الفرد i فيه (بدلا من متوسطها) بالرمز T_{2i} .

ثم حسبنا معامل الارتباط بين عمودي المجاميع T_{1i} وT_{2i} (أو المتوسطات) من العلاقة (2−1) فوجدنا أنه يساوي: $r = 0.540$.

وعندما حسبنا معامل (سبيرمان− براون) وجدنا أن قيمته تساوي : $BC = \frac{2(0.540)}{1.540} = 0.701$

وهي قيمة مقبولة وتدل على درجة ثبات مقبولة لإجابات ذلك الاستبيان .

5−2−3: أسلوب جوثمان (Guthman) :

ويعتمد هذا الأسلوب على أسلوب التجزئة بالمناصفة السابقة لأسئلة الاستبيان إلى نصفين متساويين، ثم القيام بحساب مجموع الإجابات في كل منهما فنحصل على T_{1i} وT_{2i}، ثم نقوم بحساب تباين مجاميع

الإجابات T_{1i} و T_{2i} في كل نصف على حدة فنحصل على التباينين S_1^2 و S_2^2 . ثم نقوم بحساب المجموع الكلي للإجابات مقابل كل فرد فنحصل على العمود T_i ، ثم نقوم بحساب التباين الكلي S^2 لمجموع إجابات الاستبيان، وهو يساوي تباين المجاميع الكلية الواردة في العمود T_i في الجدول (2−2)، ثم نقوم بحساب قيمة معامل الثبات للاستبيان ككل من العلاقة التي عرفها (جوثمان) التالية :

$$\propto = 2\left[1 - \frac{S_1^2 + S_2^2}{S^2}\right] \qquad (4-5)$$

ومن بيانات الجدول (2−2) واعتماداً على التجزئة السابقة نجد أن:

$$\propto = 2\left[1 - \frac{(1.524)^2 + (2.160)^2}{(3.247)^2}\right] = 0.677$$

وهي قيمة شبه مقبولة، وتدل على أن درجة ثبات الإجابات في ذلك الاستبيان يمكن أن تكون مقبولة إذا ازداد حجم العينة n، كما يمكن تحسينها بإعادة التجزئة بأسلوب آخر أو بحذف سؤال واحد أو أكثر من أسئلة الاستبيان، وسنتعرض لتلك العمليات لاحقاً.

ملاحظة: نلاحظ أن قيمة معامل (جوثمان) تختلف عن قيمة معامل (سبيرمان− براون). وذلك لأمور تتعلق بأسلوب الحساب وهنا يبرز أمامنا السؤال الثاني: أي القيمتين نعتمدها في التحليل؟

إن الجواب على هذا السؤال يعتمد على قيمتي تبايني المجموعتين S_1^2 و S_2^2 .

فإذا كان هذان التباينان متجانسين فإننا نعتمد على قيمة معامل (سبيرمان− براون).

أما إذا كان هذان التباينان مختلفين فإننا نعتمد على قيمة معامل (جوثمان).

وفي مثالنا نجد أن التباينين مختلفان لذلك نعتمد على قيمة معامل (جوثمان) ونعمل على تحسينه.

5−2−4: أسلوب ألفا كرونباخ (Cronbach's alpha):

وهو تعميم لطريقة (جوثمان) من أجل تطبيقها على جميع الأسئلة في الاستبيان المؤلف من K سؤالاً، لذلك قام (كرونباخ) باعتبار كل سؤال في الاستبيان وكأنه جزء خاص من أصل K جزءاً, واستفاد من علاقة (جوثمان)، وقام بإجراء التعميم على K جزءاً أو سؤالاً فتوصل إلى تعريف معامل جديد يسمى معامل (ألفا كرونباخ) ويحسب من العلاقة التالية:

$$\propto = \frac{k}{k-1}\left[1 - \frac{S_1^2 + S_2^2 + S_3^2 + \ldots.\ S_k^2}{S^2}\right] \qquad (5-5)$$

وإذا قمنا بتطبيق هذه العلاقة على جميع أسئلة الاستبيان نجد من الجدول (5−2) أن:

$$\propto = \frac{6}{6-1}\left[1 - \frac{(0.7379)^2 + (0.7379)^2 + (0.8756)^2 + (0.9944)^2 + (0.9429)^2 + (0.8756)^2}{(3.247)^2}\right]$$

$$\propto = \frac{6}{5}\left[1 - \frac{4.50023}{10.543000}\right] = 0.6877$$

وهو معامل الثبات لإجمالي الاستبيان، وقيمته شبه مقبولة ويمكن تحسينها بحذف سؤال أو أكثر من الاستبيان كما سنرى لاحقاً. أو بزيادة حجم العينة n .

5-2-5: كيفية استخراج الصيغ المختلفة لمعامل (ألفا كرونباخ):

يوجد لمعامل (ألفا كرونباخ) عدة صيغ رياضية هي:

أ- **صيغة التباينات المشتركة** : وتستخرج من مصفوفة التباينات المشتركة لأسئلة الاستبيان. ولنفترض أن هذه المصفوفة المتناظرة تأخذ الشكل التالي:

$$\boldsymbol{cov}(\boldsymbol{X}) = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_k \end{array} \begin{array}{c} \begin{array}{ccccc} X_1 & X_2 & X_3 & \dots & X_k \end{array} \\ \begin{bmatrix} C_{11} & C_{12} & C_{13} & \dots & C_{1k} \\ C_{21} & C_{22} & C_{23} & \dots & C_{2k} \\ C_{31} & C_{32} & C_{33} & \dots & C_{3k} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ C_{k1} & C_{k2} & C_{k3} & \dots & C_{kk} \end{bmatrix} \end{array} \qquad (\mathbf{6-5})$$

حيث رمزنا بـ C_{ij} للتباين المشترك للسؤال i مع السؤال $j(i \neq j)$، والذي يعرف بالرمز $cov(X_i, X_j)$ ، أما إذا كان $(i = j)$ فإن C_{ii} هو تباين السؤال X_i نفسه، والذي يعرف بالرمز σ_i^2 .

واعتماداً على هذه المصفوفة يمكننا أن نعرف المجموع الكلي لعناصرها، ونرمز له بـ S ونحسبه من العلاقة:

$$S = \sum_{i=1}^{k} \sum_{j=1}^{k} C_{ij} \qquad (\mathbf{7-5})$$

ثم نعرف متوسط عناصرها $\bar{S}$ ونحسبه من بالعلاقة:

$$\bar{S} = \frac{S}{k^2} = \frac{\sum \sum C_{ij}}{k^2} \qquad (8-5)$$

وحتى نبرز التأثيرات المختلفة للأسئلة على بعضها البعض، نأخذ مجموع التباينات المشتركة لها (غير القطرية) ونرمز له بـ C وهو يساوي:

$$C = \sum_{i \neq j}^{k} \sum_{j=1}^{k} C_{ij} \qquad (i \neq j) \quad (\text{غير القطرية}) \qquad (\mathbf{9-5})$$

وهكذا نجد أن:

$$S = C + \sum_{i=1}^{k} \sigma_i^2 \qquad : (\text{غير القطرية + القطرية}) \qquad (\mathbf{10-5})$$

حيث أن: $\sum_{i=1}^{k} \sigma_i^2$ هو مجموع العناصر القطرية في المصفوفة (2-6) لأن $C_{ii} = \sigma_i^2$.

إن المجموع المزدوج في (5-9) هو عبارة عن مجموع العناصر غير القطرية للمصفوفة cov والتي عددها يساوي $k(k-1)$, لذلك فإن متوسط عناصرها يساوي:

$$\bar{C} = \frac{1}{k(k-1)} \sum_{i \neq j}^{k} \sum_{j=1}^{k} C_{ij} \qquad (11-5)$$

وبناء على ذلك تم تعريف معامل (ألفا كرونباخ) الأساسي وسنرمز له بـ alpha من خلال العلاقة التالية:

$$alpha = \frac{\bar{C}}{\bar{S}} = \frac{\frac{1}{k(k-1)} \sum_{i \neq j}^{k} \sum_{j=1}^{k} C_{ij}}{\frac{1}{k^2} \sum_{i=1}^{k} \sum_{j=1}^{k} C_{ij}} = \frac{k}{k-1} * \frac{C}{S} \qquad (12-5)$$

وبما أنه لدينا من (5−10) أن C تساوي المجموع الكلي S مطروحاً منه مجموع العناصر القطرية $(\sum_{i=1}^{k} \sigma_i^2)$, أي أن:

$$C = S - \sum_{i=1}^{k} \sigma_i^2 \qquad (a\ 12-5)$$

وبذلك نجد أن العلاقة (5−12) تأخذ الشكل التالي:

$$alpha = \frac{k}{k-1} * \left[\frac{S - \sum_{i=1}^{k} \sigma_i^2}{S} \right]$$

وبعد الإصلاح نجد أن:

$$alpha = \frac{k}{k-1} \left[1 - \frac{\sum_{i=1}^{k} \sigma_i^2}{S} \right] = \frac{k}{k-1} \left[1 - \frac{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_k^2}{S} \right] \qquad (\mathbf{13-5})$$

حيث أن: K هو عدد الأسئلة و σ_i^2 هو تباين السؤال i ، وأن S هو مجموع عناصر مصفوفة التباينات المشتركة لجميع الأسئلة . وهذه الصيغة هي الأكثر انتشاراً في التطبيقات العملية، وهي تتطابق مع العلاقة السابقة (5−5), رغم اختلاف الرموز بينهما.

ب− **<u>صيغة المتوسطات</u>** : يمكن استخراج هذه الصيغة من العلاقة (5−12)، مع ملاحظة أنه لدينا من (5−10) أن: $S = \sum \sigma_i^2 + C$, فإننا نعالج العلاقة (5−12)على الشكل التالي (نضرب ونقسم البسط بـ K ثم بـ$(k-1)$):

$$alpha = \frac{k}{k-1} * \frac{C}{S} = \frac{\frac{C}{k-1}}{\frac{S}{k}} = \frac{\frac{k * C}{k(k-1)}}{\frac{\sum \sigma_i^2}{k} + \frac{C}{k}}$$

$$alpha = \frac{k * \bar{C}}{\bar{\sigma}^2 + \frac{(k-1)C}{k(k-1)}} = \frac{k\bar{C}}{\bar{\sigma}^2 + (k-1)\bar{C}} \qquad (14-5)$$

حيث أن $\bar{\sigma}^2$ هو متوسط التباينات القطرية للأسئلة المنفردة في المصفوفة (5−6).

وأن $\bar{C}$ هو متوسط التباينات المشتركة، أي متوسط العناصر غير القطرية في المصفوفة (5−6).

ج− **<u>الصيغة الارتباطية أو المعيارية:</u>** وتعتمد هذه الصيغة على عناصر المصفوفة الارتباطية بين جميع الأسئلة X_i والتي يمكن كتابتها كما يلي:

$$\boldsymbol{R} = \begin{bmatrix} 1 & r_{12} & r_{13} & \dots & r_{1k} \\ r_{21} & 1 & r_{23} & \dots & r_{2k} \\ r_{31} & r_{32} & 1 & \dots & r_{3k} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ r_{\mathrm{k}1} & r_{k2} & r_{k3} & \dots & 1 \end{bmatrix} \qquad (15-5)$$

ومنها يمكننا استنباط تعريف آخر للمعامل (ألفا كرونباخ) مشابه تماماً للتعريف الذي في (5–14)، وذلك باستخدام المصفوفة الارتباطية السابقة وباستبدال الرموز في (5–14) برموز مناسبة لمعاملات الارتباط , فنحصل على أن:

$$alpha = \frac{k\bar{r}}{1 + (k-1)\bar{r}} \qquad (16-5)$$

حيث أن $\bar{r}$ هو متوسط معاملات الارتباط غير القطرية، وأما متوسط المعاملات القطرية فيساوي (1) لأن مجموعها K، وعددها K واحداً.

ومن جهة أخرى وبناء على العلاقة (5–13) يمكننا أيضاً أن نستنتج علاقة أخرى لحساب (ألفا كرونباخ) من عناصر المصفوفة الارتباطية R مباشرة كما يلي:

$$alpha = \frac{k}{k-1}\left[1 - \frac{k}{\sum_{i=1}^{k}\sum_{j=1}^{k} r_{ij}}\right] \qquad (17-5)$$

وذلك لأن مجموع العناصر القطرية يساوي K.

وعند حساب المصفوفة R وحساب قيمة $alpha$ للمثال (2–1) نجد أن: $alpha = 0.69400$، وهي قيمة مقبولة وتدل على ثبات مقبول للإجابات في الاستبيان السابق.

د– **صيغة تحليل التباين باتجاهين:** من المعلوم أن تحليل التباين باتجاهين (two way) يعتمد على عاملين مستقلين (الأسئلة والأشخاص) ليختبر تأثيرهما على عامل ثالث (الدرجات)، وهو يعطينا جدولاً لمربعات الانحرافات حسب الأسطر وحسب الأعمدة بالإضافة لتباين الأخطاء، كما في الجدول التالي: **(انظر الفصل السابع)**

جدول (5–3): تحليل التباين الثنائي (لـ K سؤالاً من عينة حجمها n فرداً)

مصدر التباين	مجموع مربعات الانحرافات	درجة الحرية	متوسط المربعات	قيمة F	قيمة P Sig
حسب الأعمدة: column	SSC	$k-1$	$MSSC = \frac{SSC}{k-1}$	$F_C = \frac{MSC}{MSE}$	—
حسب الأسطر: raw	SSR	$n-1$	$MSSR = \frac{SSR}{n-1}$	$F_r = \frac{MSR}{MSE}$	—
الأخطاء: Error	SSE	$(n-1)(k-1)$	$MSSE = \frac{SSE}{(n-1)(k-1)}$	—	—
المجموع: Total	SST	$nk-1$	—	—	—

ومنه يمكننا حساب قيمة معامل (ألفا كرونباخ) من العلاقة :

$$alpha = \frac{MSE}{MSR} = \frac{\text{متوسط مربعات الخطأ}}{\text{متوسط مربعات الأسطر}} = \frac{1}{F_r} \qquad (18-5)$$

ملاحظة: عند حساب قيم (ألفا كرونباخ) للمثال نفسه من هذه الصيغ المختلفة ، قد نجد بعض الاختلاف بينها، وذلك يعود لعوامل التقدير وطرق الحساب والتقريب.

5–2–6: قواعد تصنيف قيم المعامل (ألفا كرونباخ) :

لقد استقرت الآراء في أغلب المراجع على تصنيف قيم (ألفا كرونباخ) التي تقع في المجال [0 , 1] إلى عدة مستويات كما في الجدول التالي:

جدول (5–4): مستويات تصنيف قيم ألفا كرونباخ

فئات التصنيف لـ $\propto$	تقدير الثبات أو الاتساق الداخلي
إذا كانت: $alpha \geq 0.90$	ممتاز
$0.80 \leq alpha < 0.90$	جيد
$0.70 \leq alpha < 0.80$	مقبول
$0.60 \leq alpha < 0.70$	هناك تساؤل
$0.50 \leq alpha < 0.60$	ضعيف
وعندما: $alpha \leq 0.50$	غير مقبول

ملاحظة: إذا كانت قيمة (ألفا كرونباخ) صغيرة أوسالبة لأحد المحاور فهذا يدل عدم ثبات الإجابات فيه، وغندها يجب العمل على التخلص من هذه الحالة بحذف واحد أو أكثر من أسئلة ذلك المحور كما سنرى لاحقاً.

5–2–7: اختبار معنوية قيمة (ألفا كرونباخ) :

لاختبار معنوية $alpha$ في المجتمع من خلال القيمة المحسوبة لـ $alpha$ ضمن مستوى دلالة $\propto=$ 0.05 , نضع الفرضيتين بناء على الجدول (5–4) كما يلي:

وهو الحل الفاصل بين القبول والرفض:

$$H_0: alpha \leq 0.70$$
$$H_1: alpha > 0.70$$

ثم نقوم بحساب مؤشر الاختبار F المعرف كما يلي:

$$F = \frac{1 - \propto_0}{1 - alpha} = \frac{1 - 0.70}{1 - alpha} \qquad (19-5)$$

وهو متحول يخضع للتوزيع $F(v_1, v_2)$، حيث أن درجتي الحرية تساويان: $v_1 = n - 1$، و$v_2 = (n-1)(k-1)$. وكتطبيق على ذلك نحسب قيمة F من (5– 19) لقيمة (ألفا) المحسوبة

من العلاقة (5−5) للمثال (5−1)، فنجد أن $F = 0.9606$، ثم نقارن هذه القيمة مع القيمة الحرجة $F_{\propto(9.45)}$ المساوية لـ $F_{\propto} = 2.096$ ونتخذ القرار كمايلي:

بما أن $F \leq F_{\propto}$ نقبل فرضية العدم التي تقول أن قيمة $alpha$ أصغر أو تساوي 0.70 .

أما إذا كانت $F > F_{\propto}$ نرفض H_0 ونقبل H_1 التي تقول أن القيمة المحسوبة لـ $alpha$ معنوية وقيمتها أكبر من 0.70 .

5−2−8 : كيفية حذف الأسئلة السيئة :

بعد أن نقوم بحساب قيمة $alpha$ واختبارها، فقد نجد أن قيمتها ضعيفة أو غير مقبولة أو مقبولة فقط، وفي مثل هذه الحالات علينا العمل على رفع قيمة $alpha$ المحسوبة، وذلك عن طريق حذف بعض الأسئلة السيئة من الاستبيان، ولإجراء ذلك الحذف علينا أن نحدد الأسئلة التي يجب حذفها. لذلك نقوم بدراسة قيم متوسطات وتباينات الأسئلة. ونركز على الأسئلة ذات المتوسطات الكبيرة أو الصغيرة، أو ذات التباينات الكبيرة، ونعمل على حذفها واحداً بعد الآخر أو دفعة واحدة، ثم نقوم بحساب قيمة $alpha$ بعد حذف كل سؤال سيئ، فإذا ازدادت قيمتها بشكل ملموس فإن ذلك يستوجب حذف ذلك السؤال من الاستبيان.

وهناك عملية خاصة في البرنامج الحاسوبي SPSS تسمى (Scale if item Deleted) (الدرجة إذا حُذف السؤال) لتنفيذ هذه العمليات وحساب تأثير حذف الأسئلة واحداً بعد الآخر على قيمة $alpha$، حيث يقوم بحساب معامل الارتباط لها مع عمود المجموع الإجمالي T، ويتم حذف كل سؤال يقابله قيمة صغيرة أو سالبة لذلك المعامل، مع الأخذ بعين الاعتبار قيمة $alpha$ بعد الحذف في كل حالة. وكمثال على ذلك نأخذ المثال التالي:

مثال (5−2): لحساب قيمة (ألفا كرونباخ) لكامل الاستبيان السابق من العلاقة (5−13) علينا أن نحسب مصفوفة التباينات المشتركة cov، لذلك نلجأ إلى الحاسوب وإلى برنامج SPSS فنجد أنها تساوي:

$$\boldsymbol{cov(X)} = \begin{matrix} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{matrix} \begin{bmatrix} X_1 & X_2 & X_3 & X_4 & X_5 & X_6 \\ 0.544 & 0.122 & -0.011 & 0.211 & 0.111 & 0.344 \\ 0.122 & 0.544 & 0.122 & 0.233 & 0.222 & 0.433 \\ -0.011 & 0.122 & 0.767 & -0.122 & -0.111 & 0.456 \\ 0.211 & 0.233 & -0.122 & 0.989 & 0.667 & 0.233 \\ 0.111 & 0.222 & -0.111 & 0.667 & 0.889 & 0.111 \\ 0.344 & 0.433 & 0.456 & 0.233 & 0.111 & 0.767 \end{bmatrix}$$

ومنها نجد أن مجموع عناصرها يساوي:

$$S = \sum_{i=1}^{k} \sum_{j=1}^{k} C_{ij} = 10.542$$

وأن مجموع العناصر القطرية فيها يساوي:

$$\sum_{i=1}^{k} \sigma_i^2 = \sum C_{ii} = 4.5$$

ومن العلاقة (5−13) نجد مباشرة أن:

$$alpha = \frac{6}{5}\left[1 - \frac{4.5}{10.542}\right] = 0.6878$$

ويمكن حساب هذا المعامل من المصفوفة الارتباطية (5−15) والعلاقة (5−17) لذلك نحسب المصفوفة الارتباطية لجميع أسئلة الاستبيان فنجد أنها تساوي:

$$R = \begin{bmatrix} 1 & 0.224 & -0.017 & 0.288 & 0.160 & 0.533 \\ 0.224 & 1 & 0.189 & 0.318 & 0.319 & 0.671 \\ -0.017 & 0.189 & 1 & -0.140 & -0.135 & 0.594 \\ 0.288 & 0.318 & -0.140 & 1 & 0.711 & 0.268 \\ 0.160 & 0.319 & -0.135 & 0.711 & 1 & 0.135 \\ 0.533 & 0.671 & 0.594 & 0.268 & 0.135 & 1 \end{bmatrix}$$

ولتطبيق العلاقة المعيارية (5−17) نقوم بحساب مجموع عناصر المصفوفة R فنجد أنه يساوي:

$$\sum_{i=1}^{6}\sum_{j=1}^{6} r_{ij} = 14.236$$

وهكذا نجد أن:

$$alpha = \frac{6}{5}\left[1 - \frac{6}{14.236}\right] = 0.694 \approx 0.70$$

وهي القيمة المعيارية لـ(ألفا كرونباخ) وهي تدل على درجة شبه مقبولة لثبات إجابات الاستبيان المدروس. ولتحسين هذه القيمة نستعين ببرنامج SPSS وندخل من Scale على (Reliability Analysis) ثم على Statistics فنجد أمامنا الخيار (Scale if item deleted)، وهو يعطينا قيمة (alpha) فيما إذا تم حذف سؤال معين من الاستبيان، ويقدم لنا بعد حسابات معقدة جدولاً مفصلاً يشير إلى قيمة (alpha) إذا تم حذف كل سؤال من أسئلة الاستبيان. وعلينا اتخاذ القرار المناسب لحذف سؤال أو أكثر لزيادة قيمة معامل الثبات.

وفي مثالنا هذا نجد أنه يعطينا الجدول التالي:

جدول (5−5): مؤشرات حذف المتحولات

المتحول المرشح للحذف	متوسط المجموع بعد الحذف	تباين المجموع بعد الحذف	معامل الارتباط المصحح للمجموع	مربع الارتباط المتعدد	قيمة alpha بعد الحذف
X_1	15.00	8.444	0.363	0.583	0.664
X_2	15.20	7.733	0.552	0.657	0.611
X_3	15.00	9.111	0.128	0.681	0.738*
X_4	15.00	7.111	0.461	0.584	0.633
X_5	15.10	7.656	0.383	0.550	0.660
X_6	15.20	6.622	0.700	0.868	0.545

وأكثر ما يهمنا في هذا الجدول هو قيمة (alpha) بعد الحذف. فنجد أن هذه القيمة تزداد لتبلغ (0.738) إذا تم حذف المتحول X_3 . وهي تتناقص إذا تم حذف أي سؤال آخر. لذلك نقوم بحذف X_3 من الاستبيان ونعيد الحسابات فنحصل على أن:

عدد المتحولات	القيمة المعيارية لـ ألفا كرونباخ	قيمة ألفا كرونباخ
5	0.740	0.738

وهما تحسبان من العلاقتين السابقتين (2–13) و(2–17)، ولكن على المتحولات الخمسة المتبقية: $(X_1 , X_2 \ , X_4 \ , X_5 \ , X_6)$.

ويمكننا إعادة الحسابات للكشف عما إذا كان يمكن حذف متحول آخر لزيادة قيمة (alpha) . ولكن تلك الحسابات تشير إلى أنه إذا تم حذف أي متحول آخر فإن قيمة (alpha) ستنقص عما هي عليه. لذلك نتوقف عن الحذف ونكتفي بحذف السؤال الثالث X_3 ، لكي نحصل على أن alpha= 0.738، على أمل أنها ستتحسن عندما يزداد حجم العينة n .

وللتأكد من معنوية هذه القيمة يمكننا إجراء اختبار F عليها فنضع الفرضيتين كما يلي:

$$H_0 \leq 0.70 \qquad , \qquad H_1 > 0.70$$

ثم نحسب قيمة مؤشر الاختبار F فنجد أن:

$$F = \frac{1 - 0.70}{1 - 0.738} = 1.145$$

ثم نبحث عن $F_{0.05}$ الحرجة والمقابلة لدرجتي الحرية:

$$v_1 = n - 1 = 10 - 1 = 9$$
$$v_2 = (n-1)(k-1) = 9 * 4 = 36 \qquad (\text{لأن أصبحت } k = 5)$$

فنجد أنها تساوي: $F_{(0.05),9.36} = 2.153$

وبما أن $F < F_{0.05}$ فإننا نقبل فرضية العدم التي تقول أن قيمة (alpha) أصغر أو تساوي من (0.70)، ونقول إن قيمة معامل الثبات مازالت ضعيفة ، لذلك يجب أن نعمل على زيادة قيمتها بزيادة حجم العينة n .

ملاحظة: يمكن حساب قيمة (alpha) لأي محور من محاور الاستبيان واعتباره كأنه استبيان خاص بحد ذاته. فمثلاً يمكننا أن نشكل من المتحولات الثلاثة الأولى $(X_1 , X_2 \ , X_3)$ محوراً خاصاً ونعتبره استبياناً كاملاً، ونحسب قيمة (alpha) له من العلاقة (2–5) فنجد أنها تأخذ الشكل التالي:

$$alpha = \frac{K}{K-1}\left[1 - \frac{S_1^2 + S_2^2 + S_3^2}{S_{T_1}^2}\right]$$

$$alpha = \frac{3}{2}\left[1 - \frac{(0.7379)^2 + (0.7379)^2 + (0.8756)^2}{(1.524)^2}\right]$$

$$alpha = \frac{3}{2}\left[1 - \frac{1.855668}{2.322576}\right] = 0.3015$$

وهي قيمة ضعيفة لثبات الإجابات في أسئلة المحور الأول (ربما بسبب X_3). ثم نشكل محوراً آخر من المتحولات (X_4, X_5, X_6) ونحسب قيمة (alpha) لها من العلاقة (2−5) التي تأخذ الشكل التالي:

$$alpha = \frac{K}{K-1}\left[1-\frac{S_4^2+S_5^2+S_6^2}{S_{T_2}^2}\right]$$

$$alpha = \frac{3}{2}\left[1-\frac{(0.9944)^2+(0.9429)^2+(0.8756)^2}{(2.160)^2}\right]$$

$$alpha = \frac{3}{2}\left[1-\frac{2.644567}{4.6656}\right] = 0.6479$$

وهي قيمة شبه مقبولة لثبات الإجابات في أسئلة المحور الثاني.

5−3: أساليب قياس الصدق :

إن مؤشرات الصدق في الاستبيان أمر هام لأنها تغني عن مؤشرات الثبات. فالإجابات الصادقة في الاستبيان هي إجابات ثابتة، ولكن العكس غير صحيح، لأنه ليست بالضرورة أن تكون الإجابات الثابتة صادقة (فقد تكون متحيزة وبالتالي تكون غير صادقة).

والصدق: هو التطابق أو التوافق بين الإجابات التجريبية في الاستبيان مع القيم المعيارية أو الفعلية لها، إذاً لقياس الصدق لابد من وجود قيم معيارية معلومة لنقيس عليها الإجابات في كل سؤال. مثل: ماهو تقديرك لدرجة الحرارة الآن دون أن تعلم أنها على المقياس تساوي 25° ، وهنا تؤخذ القيمة (25) معياراً لصدق الإجابات.

وللصدق أنواع وأشكال هي:

- الصدق الظاهري أو الشكلي: وهو التطابق أو التوافق بين فقرات الاستبيان من حيث الشكل ومن حيث الغرض الذي نقيسه.
- صدق المحتوى أو المضمون: وهو أن يشمل الاستبيان جميع جوانب الموضوع المدروس ويعبر عن مضامينه.
- صدق المفهوم: وضوح العبارات والأسئلة ومطابقتها مع المفهوم العام للفكرة المطروحة (إن زيادة الدراسة يزيد من معدل النجاح).
- الصدق العاملي: وهو كيفية تحليل الصفة المدروسة إلى عناصرها الأولية لتسهيل عملية قياسها.
- الصدق التنبؤي: يقصد به التطابق أو الارتباط لفكرة معينة حالية مع فكرة أخرى في المستقبل (تفوق الطالب في الثانوية يؤدي إلى تفوقه في الجامعة) ويتم معالجة هذه الأنواع السابقة من قبل المختصين والمحكمين.
- الصدق التلازمي: ويقصد به تطابق الإجابات أو متوسطاتها الأفقية مع متوسطات أو مع إجابات قياسية تسمى المعيار أو (المحك)، والمحك هو مقياس خارجي لا يتعلق بالاختبار، وهو مقياس

موضوعي، سبق إن تم التأكد من صدقه وثباته (مثل: متوسط ضغط الدم الطبيعي= 120 بار، أو متوسط درجة حرارة جسم الإنسان= 36.5 درجة). ويتم التأكد من الصدق التلازمي من خلال دراسة الارتباط بين إجابات الاستبيان لكل سؤال مع درجات المحك المقابلة لها. وكلما كان الارتباط شديداً بينهما كان الصدق التلازمي في الاختبار محققاً.

- الصدق التمييزي: وهو يعبر عن قدرة الاستبيان على تحديد التطابق بين طرفي الإجابات في المجموع العام أو في المتوسط العام لذلك الاستبيان.

وهناك عدة مؤشرات لقياس الصدق ونذكر منها التالي:

5-3-1: معامل الارتباط مع المحك: وهو يستخدم لقياس الصدق التلازمي لإجابات كل سؤال مع المحك المعتمد، ويحسب من العلاقة (2-1) .

ونظراً لعدم وجود محكات جاهزة لأسئلة الاستبيان، لذلك نأخذ المتوسطات العامة لقيم الإجابات حسب الأفراد $\bar{x}_i$ ونشكل منها عموداً خاصاً. ونعتبره المحك المعياري لصدق الإجابات في ذلك الاستبيان.

ولإنشاء محك لمثالنا السابق (5-1) (قبل حذف المتحول X_3)، قمنا بحساب متوسطات إجابات الأفراد على جميع الأسئلة ووضعناها في عمود خاص ورمزنا له بـ $\bar{x}_i$ (لكل فرد i). ثم قمنا بحساب معاملات الارتباط بين عمود الإجابات في كل سؤال مع ذلك المحك $\bar{x}_i$ فحصلنا على الجدول التالي:

المتحولات	X_1	X_2	X_3	X_4	X_5	X_6
معامل الارتباط مع المحك $\bar{x}_i$	0.552	0.700*	0.387	0.685*	0.617	0.825**
P=Sig (2-tailed)	0.098	0.024	0.269	0.029	0.057	0.003

ومن هذا الجدول نلاحظ أن قيم معامل الارتباط معنوية فقط مع ثلاثة متحولات هي: (X_2, X_4, X_6)، وإن القيم العددية لمعاملاتها تشير إلى درجة مقبولة لصدق الإجابات لتلك الأسئلة وإن قيمه غير معنوية مع المتحولات الأخرى (X_1, X_3, X_5)، وهي تشير إلى أن درجة الصدق في إجابات هذه المتحولات ضعيفة.

وهنا نلاحظ إن قيمته مع X_3 صغيرة جداً. لذلك نشكك في الإجابات على X_3 وننصح بحذفه من الاستبيان، وهذا ما أشرنا إليه سابقاً.

وأخيراً نشير إلى أن جميع هذه المعاملات ستتحسن وتستقر عند قيم معينة لكل منها عندما نزيد حجم العينة n، ويجب أن لا نتسرع بحذف المتحولات ذات القيم الصغيرة قبل أن نستكمل تمرير الاستبيان على كامل أفراد العينة، حيث يمكن حذف أي سؤال من الاستبيان بعد ذلك.

5-3-2: **معامل الصدق العام:** ويستخدم لقياس درجة الصدق العام, للاستبيان ككل أو لكل محور من محاوره، ويحسب من الجذر التربيعي لمعامل الثبات (alpha) كما يلي:

$$V = \sqrt{alpha} \qquad (20-5)$$

وفي مثالنا (5−1) نجد أن معامل الصدق لذلك الاستبيان تساوي:

$$V = \sqrt{0.6878} = 0.8293$$

وهي تعكس درجة جيدة لمصداقية الإجابات في ذلك الاستبيان.

ولحساب معامل الصدق للمحور الأول الذي يشمل (X_1, X_2, X_3) نجد أن:

$$V_1 = \sqrt{0.3015} = 0.5490$$

وهي تعكس درجة ضعيفة لمصداقية الإجابات في ذلك المحور، وربما بسبب الإجابات في X_3 المرفوض.

ولحساب معامل الصدق للمحور الثاني الذي يشمل (X_4, X_5, X_6) نجد أن:

$$V_2 = \sqrt{0.6497} = 0.8060$$

وهي تعكس درجة جيدة لمصداقية الإجابات في ذلك المحور .

5−3−3: اختبار Z أو t : وهو يستخدم لاختبار صدق إجابات كل سؤال j ، بمقارنة متوسطه العمودي $\bar{x}_j$ المسجل في السطر الأخير مع المتوسط المتوقع له. ولتحديد المتوسط المتوقع له نلجأ إلى اعتماد القيمة الوسطى لخيارات الأجوبة (مثل العدد 3 في المقياس الخماسي)، كما يمكن اعتبار المتوسط العام $\bar{\bar{x}}$ كقيمة متوقعة لجميع الإجابات في الاستبيان. ثم نضع الفرضيتين كما يلي:

$$H_0: \bar{x}_j = \bar{\bar{x}} = 3 \quad , \qquad H_1: \bar{x}_j \neq \bar{\bar{x}}$$

وبعد ذلك يتم حساب قيمة مؤشر الاختبار t لكل سؤال j ، من العلاقة:

$$t_j = \frac{\bar{x}_j - \bar{\bar{x}}}{S_j/\sqrt{n}} \approx \frac{\bar{x}_j - 3}{S_j/\sqrt{n}} \qquad (21-5)$$

ثم نقارن قيمة t مع القيمة الحرجة $t_{\frac{\alpha}{2}}$ عند $(n-1)$ درجة حرية ، فإذا كانت $|t| < t_{\frac{\alpha}{2}}$ نقبل فرضية العدم التي تقول بعدم وجود فرق معنوي لمتوسط إجابات السؤال j مع القيمة المتوقعة لها.

وعند تطبيق ذلك على السؤال الأول في مثالنا السابق نجد من بيانات الجدول (2−2) أن: (باعتماد المتوسط العام) .

$$t_1 = \frac{\bar{x}_1 - \bar{\bar{x}}}{S_1/\sqrt{n}} = \frac{3.1 - 3.017}{0.7379/\sqrt{10}} = 0.3557$$

ومن جداول توزيع (ستودينت) نجد أن القيمة الحرجة المقابلة لمستوى دلالة $\alpha = 0.05$ ولدرجة حرية $(n-1=9)$ تساوي $t_{\frac{\alpha}{2}*9} = 1.833$ ، وعند المقارنة نجد أن $t_1 < t_{\frac{\alpha}{2}}$, لذلك نقبل فرضية العدم H_0 ونعتبر أن إجابات السؤال الأول صادقة باحتمال 0.95 . وهكذا يمكننا اختبار صدق الإجابات في الأسئلة الأخرى.

ملاحظة: إذا كان الاستبيان مؤلفاً من عدة محاور. فيمكن دراسة واختبار صدق متوسطات كل محور بمقارنتها مع المتوسط العام للاستبيان، وتطبيق نفس العلاقة (5−21) على متوسطات كل محور كما فعلنا أعلاه .

5−3−4: اختبار الصدق التمييزي: ويستخدم لاختبار تطابق طرفي الإجابات، ويشترط هذا الاختبار أن يكون حجم العينة n كبيراً، وهو يعتمد على مقارنة طرفي الإجابات المرتبة لكل سؤال أو للاستبيان ككل. وهذا يقتضي تشكيل مجموعتين طرفيتين من إجابات السؤال أو من المتوسطات العامة للإجابات في الاستبيان ككل. وسنقوم باختصار خطوات العمل لاختبار الصدق التمييزي للاستبيان ككل بما يلي:

أ− نرتب قيم المتوسطات العامة للاستبيان $\bar{x}_i$ تصاعدياً (أو تنازلياً) ونضعها في عمود جديد.

ب− نأخذ حوالي 25% أو 30% من المتوسطات العامة المرتبة الأولى، ونشكل منها المجموعة الطرفية الأولى ونضعها في عمود جديد آخر، ثم نحسب متوسطها $\bar{x}_1$ وتباينها S_1^2 .

ج− نأخذ حوالي 25% أو 30% من المتوسطات العامة المرتبة الأخيرة، ونشكل منها المجموعة الطرفية الثانية ونضعها في عمود جديد ثالث مقابل عمود المجموعة الأولى، ثم نحسب متوسطها $\bar{x}_2$ وتباينها S_2^2 ، ويفضل أن يكون عدد العناصر في المجموعتين متساوياً.

د− نضع الفرضيتين حول الفرق بين هذين المتوسطين في المجتمع كما يلي:

$$H_0: \bar{y}_1 - \bar{y}_2 = 0 \qquad\qquad H_1: \bar{y}_1 - y_2 \neq 0$$

هـ− نقوم بحساب مؤشر الاختبار t للفرق بين عينتين مستقلتين من العلاقة:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \qquad (22-5)$$

و− نقوم بمقارنة قيمة t المحسوبة مع قيمة $t_{\frac{\propto}{2}}$ الحرجة والمقابلة لدرجة حرية المساوية لأصغر العددين $(n_1 - 1)$ أو $(n_2 - 1)$ حسبما ورد في العلاقة (1−21)، فإذا كانت $|t| < t_{\frac{\propto}{2}}$ نقبل فرضية العدم ونعتبر أن الإجابات على ذلك الاستبيان صادقة باحتمال 0.95 . والعكس بالعكس.

الفصل السادس
اختبارات الفرضيات البسيطة

6-1: تمهيد :

تعتمد اختبارات الفرضيات على البيانات المتوفرة عن الظواهر المدروسة في المجتمعات الإحصائية. والبيانات هي قيم عددية أو حالات وصفية تعبر عن المتحولات المعرفة على الظاهرة المدروسة، وبذلك ويمكننا تصنيف البيانات إلى نوعين أساسيين هما:

أ- بيانات كمية: وهي بيانات عددية عن متحولات قابلة للقياس بواحدات قياس محددة، وهذه البيانات يمكن أن تكون:

- منقطعة: كعدد أفراد الأسرة- وعدد الطلاب- وعدد السيارات ...الخ.
- مستمرة: كعمر الإنسان- درجة الحرارة- مقدار الدخل ....الخ.

ب- بيانات نوعية: وهي حالات وصفية لمتحولات غير قابلة للقياس، وهذه المعلومات يمكن أن تكون:

- أسمية: كحالات الجنس- حالات العمل- الحالة الاجتماعية ...الخ.
- مرتبة: كحالات التعليم- حالات الوظيفة- حالات الرضا ..الخ.

ويتم تجميع البيانات عن الظاهرة المدروسة أو عن المتحولات المطلوبة من عناصر المجتمع الاحصائي بواسطة أحد الأسلوبين :

- الحصر الشامل: وهو يشمل جميع عناصر المجتمع الاحصائي المؤلف من N عنصراً .
- المسح بالعينة: وهو يشمل جزء من المجتمع ويكون على شكل عينة حجمها n عنصراً, تسحب عشوائياً من عناصر ذلك المجتمع بدون إعادة أو مع الاعادة.

وتستخدم بيانات هذه العينة لتقدير معالم المجتمع المجهولة مثل: المتوسط μ أو التباين σ^2 أو النسبة فيه R، وذلك من خلال استخدام المؤشرات المقابلة لها في العينة، والتي سنسميها (مؤشرات العينة)، وهي متوسط العينة $\bar{x}$ وتباين العينة S^2 والنسبة في العينة r . ويبُرهن في نظرية العينات أن مؤشرات العينة المصححة هي تقديرات غير متحيزة وفعالة ومتماسكة لمعالم المجتمع المقابلة لها.

والآن لنفترض أننا نقوم بدراسة خصائص أحد المتحولات الكمية X من عناصر المجتمع المدروس (وليكن X وزن الطالب في الجامعة)، لذلك نسحب عينة عشوائية من طلاب ذلك المجتمع بحجم n طالباً, فنحصل من كل طالب i فيها على وزن محدد x_i , ويكون لدينا القياسات التالية:

$$X: x_1 \quad x_2 \quad x_3 \dots\dots x_i \dots\dots x_n$$

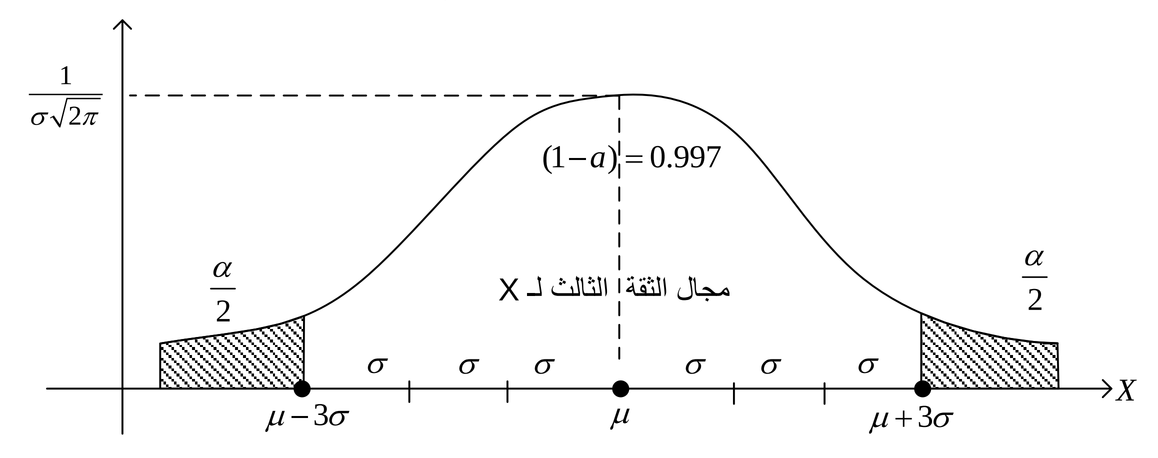
وبناء على نظرية العينات نقوم بتقدير معالم المجتمع المجهولة من مؤشرات العينة المعلومة وحساب مقدار الخطأ المرتكب في كل تقدير وفق الجدول التالي :

جدول (6−1): معالم المجتمع وتقديراتها من مؤشرات العينة :

معالم المجتمع المجهولة للمتحول	تقديرات المعالم من مؤشرات العينة
1− قيمة المتوسط في المجتمع μ	تقدر من متوسط العينة: $\bar{x} = \frac{1}{n}\sum x_i$ ونكتب ذلك كما يلي: $\tilde{\mu} = \bar{x}$
2− قيمة تباين المجتمع σ^2	تقدر من تباين العينة المصحح والمعرف بالعلاقة: $s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$ ونكتب ذلك كما يلي: $\tilde{\sigma}^2 = s^2$
3− قيمة النسبة في المجتمع R للمتحول الثنائي (0 او 1)	تقدر من النسبة في العينة $r = \frac{m}{n}$ ونكتب ذلك كما يلي: $\tilde{R} = r$ حيث m عدد الظهور و n حجم العينة وتباينها يساوي $s^2 \approx r(1-r)$
4− الخطأ المعياري المرتكب في تقدير متوسط المجتمع في حالة السحب مع الاعادة: $\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$	يقدر من خلال ما يقابله في العينة: $\tilde{\sigma}_{\bar{x}} = \sqrt{\frac{s^2}{n}} = \frac{s}{\sqrt{n}}$
5− الخطأ المعياري المرتكب في تقدير المتوسط في حالة السحب بدون إعادة: $\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1} * \frac{\sigma^2}{n}}$	يقدر من خلال العينة بالعلاقة: $\tilde{\sigma}_{\bar{x}} = \sqrt{\frac{N-n}{N-1} * \frac{s^2}{n}} \approx \sqrt{\frac{N-n}{N} * \frac{s^2}{n}}$
6− الخطأ المعياري المرتكب في تقدير النسبة R في حالة السحب مع الاعادة: $\sigma_r = \sqrt{\frac{R(1-R)}{n}}$	يقدر من خلال خطأ النسبة r في العينة كمايلي: $\tilde{\sigma}_r = \sqrt{\frac{r(1-r)}{n}}$
7− الخطأ المعياري المرتكب في تقدير النسبة R في حالة السحب بدون إعادة: $\sigma_r = \sqrt{\frac{N-n}{N-1} * \frac{R(1-R)}{n}}$	يقدر من خلال خطأ النسبة r في العينة كما يلي: $\tilde{\sigma}_r = \sqrt{\frac{N-n}{N-1} * \frac{r(1-r)}{n}} = \sqrt{\frac{N-n}{N} * \frac{r(1-r)}{n}}$

ولكن عملية التقدير لا تنتهي عند ذلك , بل يجب إنشاء مجال ثقة يحتوي المعلم الذي نقدره في المجتمع. وحتى نستطيع إنشاء مجال ثقة يجب أن يكون التوزيع الاحتمالي للمتحول X معلوماً. ومنه يجب أن يكون توزيع متوسط العينة $\bar{x}$ معلوماً أيضاً.

واختصاراً لهذه القضايا نفترض أن المتحول المدروس X يخضع في المجتمع للتوزيع الطبيعي العام $N(\mu , \sigma^2)$ الذي متوسطه μ وتباينه σ^2 وهو يرسم الشكل التالي :



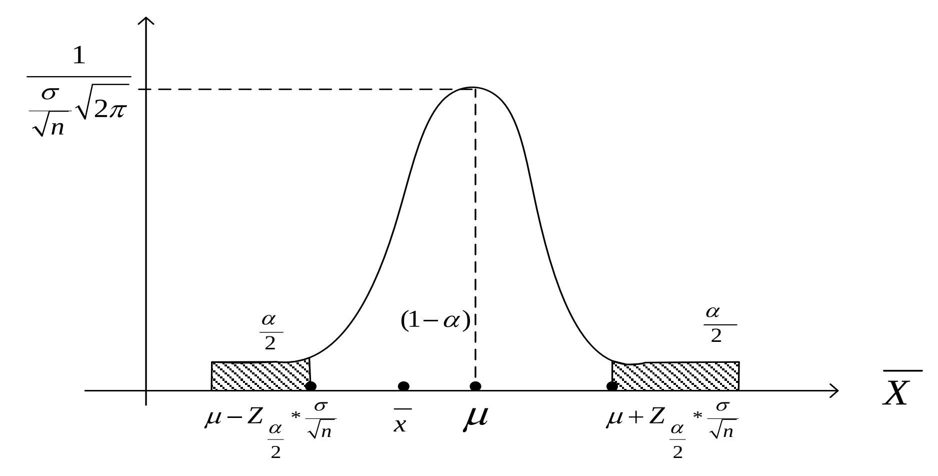
الشكل (6−1): منحنى التوزيع الطبيعي للمتحول X ومجال الثقة الثالث

وبناءً على خواص التوزيع الطبيعي يمكننا إنشاء مجال الثقة لـ X المقابل لاحتمال ثقة $(1-\propto)$ أو لمستوى دلالة $(\propto)$ بحيث يكون:

$$P\left[\mu - Z_{\frac{\propto^*}{2}}\sigma \leq X \leq \mu + Z_{\frac{\propto^*}{2}}\sigma\right] = 1-\propto \qquad (1-6)$$

حيث أن $Z_{\frac{\propto}{2}}$ هي قيمة متحول التوزيع الطبيعي المعياري التي تترك نصف مستوى الدلالة $\left(\frac{\propto}{2}\right)$ على يمينها و$\left(\frac{\propto}{2}\right)$ على يسار $-Z_{\frac{\propto}{2}}$, [حيث استبدلنا الرمز $Z_{1-\frac{\propto}{2}}$ بالرمز $Z_{\frac{\propto}{2}}$ للاختصار]. ولقد أنشأنا على الشكل (6−1) مجال الثقة الثالث الذي يقابل احتمال ثقة (0,997).

وعندما نسحب عينة من عناصر المجتمع بحجم n نحصل منها على متوسط هو $\bar{x}$ وعلى تباين هوs^2، ولكن هذه العينة ليست وحيدة بل يمكن أن يسحب غيرنا وغيرنا عينات أخرى، فيحصل على متوسطات أخرى $\bar{x}_k$ وتباينات أخرى s_k^2، وبما أن عدد العينات الممكنة يساوي C_N^n عينة (في حالة السحب بدون إعادة)، فإن هذا يعني أنه يمكننا أن نحصل على C_N^n متوسطاً $\bar{x}_k$، وكل منها يعتبر تقديراً لمتوسط المجتمع μ . وبناءً عليه يكون متوسط العينة $\bar{x}$ هو الآخر متحولاً عشوائياً جديداً متوسطه μ أيضاً، ولكن تباينه يساوي $\frac{\sigma^2}{n}$ (وانحرافه المعياري يساوي $\frac{\sigma}{\sqrt{n}}$) ويخضع للتوزيع الطبيعي العام $N\left(\mu , \frac{\sigma^2}{n}\right)$، والذي يأخذ الشكل الضامر والمتطاول التالي:



الشكل (6–2): توزيع متوسط العينة $\bar{x}$

وبناءً على شكل التوزيع (1–2) وقياساً على العلاقة (6–1) يمكننا أيضاً إنشاء مجال الثقة للمتوسط $\bar{x}$ المقابل لاحتمال الثقة (1–∝) أو لمستوى الدلالة (∝) بحيث يكون:

$$P\left[\mu - Z_{\frac{\propto}{2}} * \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu + Z_{\frac{\propto}{2}} * \frac{\sigma}{\sqrt{n}}\right] = 1-\propto \qquad (2-6)$$

وحيث أن $Z_{\frac{\propto}{2}}$: هي القيمة الجدولية (الحرجة) لمتحول التوزيع الطبيعي المعياري المقابلة لنصف مستوى الدلالة $\left(\frac{\propto}{2}\right)$ على اليمين و $\left(\frac{\propto}{2}\right)$ على اليسار، وهذا المجال يضمن لنا أن تقع قيمة $\bar{x}$ ضمنه باحتمال يساوي (1–∝).

وحتى نستفيد من العلاقة (6-2) نطرح من أطرافها μ ثم نقسمها على $\frac{\sigma}{\sqrt{n}}$ فنحصل على ما يلي:

$$P\left[-Z_{\frac{\propto}{2}} \leq \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leq Z_{\frac{\propto}{2}}\right] = 1-\propto \qquad (3-6)$$

وهذا المجال يضمن لنا أن تقع قيمة المقدار $\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)$ ضمن المجال $\left[-Z_{\frac{\propto}{2}}\ , +Z_{\frac{\propto}{2}}\right]$ باحتمال يساوي (1–∝)، وهنا نلاحظ أن المقدار $\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)$ هو متحول عشوائي ثالث، وهنا نميز بين حالتين هما:

أ– إذا كانت قيمة σ^2 وبالتالي قيمة σ معلومة من المجتمع: فإن المقدار $\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)$ يخضع للتوزيع الطبيعي المعياري $N(0\,,1)$ لأنه ناتج عن معايرة المتحول المتوسط $\bar{x}$، لذلك نرمز له بـ Z ونكتبه كما يلي:

$$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \qquad (4-6)$$

وهكذا نجد أنه يمكننا اعتبار هذا المقدار مؤشراً لاختبار أي فرضية حول متوسط المجتمع μ (مثل الفرضية $H_0: \mu = \mu_0$)، وذلك عندما نعوض μ بـ μ_0 ونقوم بحساب قيمة Z من العلاقة (6–4) وبشرط أن تكون قيمة σ معلومة.

ثم نقارن قيمة Z المحسوبة مع طرفي المجال المعرف في [6–3] وهو $\left[-Z_{\frac{\propto}{2}}\ , +Z_{\frac{\propto}{2}}\right]$. ونتخذ القرار كمايلي:

– إذا كانت قيمة Z المحسوبة واقعة ضمن ذلك المجال نقبل تلك الفرضية $(H_0: \mu = \mu_0)$ ،

– أما إذا كانت قيمة Z المحسوبة واقعة خارجه من الطرفين، فإننا نرفض الفرضية $(H_0: \mu = \mu_0)$،

وبعبارة أخرى إذا كانت $|Z| \leq Z_{\frac{\propto}{2}}$ نقبل الفرضية المذكورة ، وإذا كانت $|Z| > Z_{\frac{\propto}{2}}$ نرفض الفرضية السابقة. علماً بأن احتمال قبول الفرضية $(H_0: \mu = \mu_0)$ يساوي $(1-\propto)$ واحتمال رفضها من الطرفين يساوي $(\propto)$.

ب–أما إذا كانت قيمة σ^2 في المجتمع مجهولة, فإننا نلجأ إلى تقديرها من خلال تباين العينة المصحح s^2 , ونقوم باستبدال قيمة σ في $(4-1)$ بتقديرها s من العينة، فنحصل على متحول عشوائي جديد مركب من متحولين عشوائيين $\bar{x}$ و s ونرمز له بـ t ونكتبه كما يلي:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \qquad (5-6)$$

ومعلوم من نظرية الاحتمالات والاحصاء الرياضي أن المتحول t يخضع لتوزيع (ستودينت) ذي(n-1) درجة حرية(وهو توزيع يتقارب مع التوزيع الطبيعي المعياري عندما تصبح $n \geq 30$) . وقياساً على العلاقة (6–3) يمكننا أن ننشأ مجال الثقة للمتحول t كما يلي:

$$P\left[-t_{\frac{\propto}{2}} \leq \frac{\bar{x} - \mu}{s/\sqrt{n}} \leq +t_{\frac{\propto}{2}}\right] = 1-\propto \qquad (6-6)$$

حيث أن $t_{\frac{\propto}{2}}$ هي قيمة متحول (ستودينت) الجدولية(أوالحرجة), المقابلة لنصف مستوى الدلالة $\left(\frac{\propto}{2}\right)$ على اليمين ولـ $\left(\frac{\propto}{2}\right)$ على اليسار ولـ (n-1) درجة حرية، وهذا المجال يضمن لنا أن تقع قيمة t ضمن المجال $\left[-t_{\frac{\propto}{2}}\ ,+t_{\frac{\propto}{2}}\right]$ باحتمال قدره $(1-\propto)$. وهكذا نجد أنه يمكننا الاستفادة من المتحول t في حالة العينات الصغيرة في اختبار أي فرضية حول متوسط المجتمع μ (مثل $H_0: \mu = \mu_0$) فنعوض μ بـ μ_0 ونحسب قيمة t من العلاقة (6–5), ثم نقارنها مع طرفي المجال المعرف في (6–6) وهو المجال $\left[-t_{\frac{\propto}{2}}\ ,\ +t_{\frac{\propto}{2}}\right]$, ونتخذ القرار كما يلي:

- إذا كانت قيمة t المحسوبة واقعة ضمن ذلك المجال نقبل تلك الفرضية $(H_0: \mu = \mu_0)$،
- أما إذا كانت قيمة t المحسوبة واقعة خارجه فإننا نرفض الفرضية $(H_0: \mu = \mu_0)$.

وبعبارة أخرى: إذا كانت $|t| \leq t_{\left(\frac{\propto}{2},n-1\right)}$ فإننا نقبل الفرضية $(H_0: \mu = \mu_0)$، أما إذا كانت $|t| > t_{\left(\frac{\propto}{2},n-1\right)}$ ، فإننا نرفض الفرضية المذكورة. علماً بأن احتمال قبول الفرضية $(\mu = \mu_0)$ يساوي $(1-\propto)$ واحتمال رفضها من الطرفين يساوي $(\propto)$.

وهكذا نجد أنه يمكننا، وباتباع نفس الأسلوب، استنباط العديد من المؤشرات لاستخدامها في اختبارات الفرضيات المختلفة (كل حالة حسب طبيعتها وحسب توزيعها الاحتمالي), فنحصل على مؤشرات لتقدير النسبة R والتباين σ^2 وغيرهما.

ومن جهة أخرى يمكننا أن ننشأ مجال ثقة يحوي متوسط المجتمع μ وذلك بمعالجة العلاقة (6–2), وذلك بطرح المقدار $(\bar{x}-\mu)$ من أطرافها فنحصل على المجال التالي:

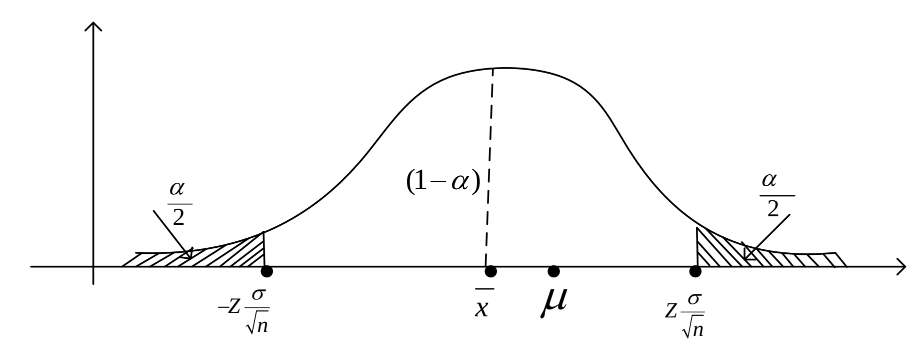
$$P\left[\bar{x}-Z_{\frac{\alpha}{2}}*\frac{\sigma}{\sqrt{n}}\leq\mu\leq\bar{x}+Z_{\frac{\alpha}{2}}*\frac{\sigma}{\sqrt{n}}\right]=1-\alpha \qquad (7-6)$$

وهو مجال مركزه متوسط العينة $\bar{x}$ (وليس μ) ونصف طوله يساوي $\left(Z_{\frac{\alpha}{2}}*\frac{\sigma}{\sqrt{n}}\right)$, ويضمن لنا أنه يحتوي على متوسط المجتمع المجهول μ باحتمال $(1-\alpha)$ والشكل (6–3) يوضح ذلك.

وإذا كان التباين σ^2 مجهولاً فإننا نستبدله بتقديره s^2، وعندها فإن مجال الثقة (6–7) يصبح معرفاً على توزيع (ستودينت) ذي $(n-1)$ درجة حرية ويأخذ الشكل التالي:

$$P\left[\bar{x}-t_{\frac{\alpha}{2}}*\frac{s}{\sqrt{n}}\leq\mu\leq\bar{x}+t_{\frac{\alpha}{2}}*\frac{s}{\sqrt{n}}\right]=1-\alpha \qquad (8-6)$$

وهو مجال مركزه $\bar{x}$ ونصف طوله $\left(t_{\frac{\alpha}{2}}*\frac{S}{\sqrt{n}}\right)$, ويضمن لنا أنه يحتوي على متوسط المجتمع المجهول μ باحتمال $(1-\alpha)$.



الشكل (6–3): مجال الثقة لـ μ

6–2: : أنواع وأسماء أهم الاختبارات :

توضع الفرضيات على معالم المجتمع (مثل μ أو R أو σ^2) أو على بعض خصائصه (مثل: المساواة، الفرق، الاستقلال، التوافق، التكرار، الالتواء ...الخ) وتختبر باستخدام معلومات العينة وبواسطة مؤشر الاختبار المناسب. وتصنف الاختبارات إلى نوعين أساسيين :

أ- الاختبارات المعلمية: وتطبق على المتحولات الكمية .

ب- الاختبارات اللامعلمية: وتطبق على المتحولات النوعية والرتبية.

كما يمكن تصنيف الاختبارات حسب عدد المجتمعات والعينات كما يلي :

1- اختبارات لمجتمع واحد (عينة واحدة) .

2- اختبارات لمجتمعين (عينتين مستقلتين) .

3- اختبارات لعدة مجتمعات (لعدة عينات مستقلة) .

4- اختبارات لعينتين مترابطتين

5- اختبارات لعدة عينات مترابطة .

ويتضمن الجدول التالي أهم الاختبارات المعلمية واللامعلمية ومجالات تطبيقها.

جدول (6-2): أهم الاختبارات الاحصائية المعلمية واللامعلمية:

أهم الاختبارات المعلمية	أهم الاختبارات اللامعلمية
1- اختبار Z الطبيعي لعينة واحدة وهو يطبق على متوسط المجتمع μ وعلى النسبة R فيه، مثل العلاقة (6-4)	**1- اختبار χ^2 : للاستقلال والارتباط بين متحولين نوعيين أو أحدهما نوعي . في عينة واحدة (غير مرتبة)**
2- اختبار (ستودينت) t لعينة واحدة صغيرة وهو يطبق على μ وعلى R مثل العلاقة (6-5)	**2- اختبار Gamma: للاستقلال والارتباط بين متحولين مرتبين من عينة واحدة**
3- اختبار χ^2 لعينة واحدة ويطبق على تباين المجتمع σ^2	**3- اختبارات الثبات والصدق ويستخدم في الاستبيانات (للمعلومات المرتبة)**
4- اختبار Z الطبيعي لعينتين مستقلتين ويطبق على الفرق بين متوسطي المجتمعين أو على الفرق بين النسبتين فيهما مثل العلاقة (6-24)	**4- اختبار مكنمارا: للحالات غير المرتبة (جدول رباعي).**
5- اختبار (ستودينت) t لعينيتين مستقلتين ويطبق على الفرق $\mu_1 - \mu_2$ وعلى الفرق $R_1 - R_2$، مثل العلاقة (6-25)	**5- اختبار: ويلكوكسن للرتب المؤشرة للمتحولات المرتبة، والمأخوذة من عينتتين مرتبطتين.**
6- اختبارF لعينتين مستقلتين ويطبق على تبايني مجتمعين $\left(\frac{\sigma_1^2}{\sigma_2^2}\right)$، مثل العلاقة (6-35)	**6- اختبار كروسكال-وايلز: للمتحولات المرتبة، والمأخوذة من عدة عينات مستقلة.**
7- اختبار t للفرق بين الأزواج المتقابلة (عينتين مترابطتين) مثل العلاقة (6-54)	**7- اختبار مان ويتني: للمتحولات المرتبة، والمأخوذة من عينتين مستقلتين.**
8- اختبار تحليل التباين الأحادي لأكثر من عينتين مستقلتين، مثل العلاقة (6-40)	**8- اختبارفريدمان: للمتحولات المرتبة، والمإخوذة من عدة عينات مستقلة.**
9- اختبار تحليل التباين الثنائي لمؤشرين على عدة مجتمعات .	**9- اختبار t_C كيندال: للمتحولات المرتبة**
10- اختبار χ^2 لتوافق التوزيعات الاحتمالية من عينة واحدة	**10- اختبار الاشارة للمتحولات الثنائية (1 , 0)**
11- اختبار كولموغوروف- سميرنوف لتوافق التوزيعات الاحتمالية	**11- اختبار كوكران- مينتال للبيانات المرتبة**
12- اختبار معنوية معامل الارتباط الخطي (معامل بيرسون)	**12- اختبار معنوية معامل الارتباط الرتبي (سبيرمان) للمتحولات المرتبة**

6−2−1: حساب موثوقية وقوة الاختبار:

في الحقيقة إن عملية إجراء اختبار لأية فرضية عدم H_0، تتأثر بعدة عوامل أهمها حقيقة الفرضية H_0 في المجتمع ونوع القرار المتخذ بشأنها، وبذلك نجد أنه يكون لدينا الحالات التالية:

- إن فرضية العدم H_0 قد تكون بحقيقتها صحيحة أو خاطئة .
- إن القرار الذي سنأخذه حول H_0 يمكن أن يكون قبولاً أو رفضاً لها .

ويمكن وضع تقاطعات هذه الحالات الأربع في جدول كالتالي:

جدول(6−3) : **حالات تقاطع حقيقة فرضية العدم مع نوع القرار المتخذ** حولها

نوع القرار المتخذ / **حقيقة الفرضية** H_0	**قبول**	**رفض**
H_0 **صحيحة**	**القرار صحيح واحتماله** $=1-\propto$	**القرار غير صحيح واحتماله** $=\propto$
H_0 **خاطئة**	**القرار غير صحيح واحتماله** $=c$	**القرار صحيح واحتماله** $= 1-c$

ومن الجدول السابق نلاحظ إنه عندما نتخذ القرار حول H_0 , فيمكن أن يكون قرارنا غير صحيح في الحالتين التاليتين: رفض الفرضية الصحيحة، قبول الفرضية الخاطئة، وعندها سنرتكب أحد الخطأين التاليين:

- خطأ النوع الأول error type I: وهو قرار رفض الفرضية H_0 رغم إنها صحيحة، وإن احتمال وقوعنا في هذا الخطأ يسمى مستوى الدلالة $\propto$، ويسمى الاحتمال المتمم له $(1-\propto)$ بدرجة الثقة أو بالموثوقية.
- خطأ النوع الثاني error type II : وهو قرار قبول الفرضية H_0 رغم إنها خاطئة أو غير صحيحة، وإن احتمال وقوعنا في هذا الخطأ يساوي عدداً آخراً c، ويسمى الاحتمال المتمم له $(1-c)$ بقوة الاختبار .

وبناءً على ذلك تم تعريف قوة الاختبار: بأنها احتمال رفض الفرضية H_0 عندما تكون خاطئة. وهو يتمم احتمال قبولها c، أي أن قوة الاختبار تعرف بالاحتمال المتمم لـ c وهو يساوي:

$$W = 1 - c \qquad (23-6)$$

ويتم حساب قيم c من تكاملات شرطية معقدة لا مجال للخوض فيها في هذا الفصل. وللتعمق في ذلك يمكن الرجوع إلى كتاب الإحصاء الرياضي للمؤلف صفحة 151, أو إلى أي مرجع مختص آخر.

6-3: اختبارات معالم مجتمع طبيعي (من عينة واحدة):

6-3-1: اختبار متوسط المجتمع μ (أو النسبة R):

ويتألف من الخطوات التالية:

1- نحدد مستوى الدلالة $(\propto)$ أو احتمال الثقة $(1-\propto)$, وعادة يتم وضعه $(\propto = 0.05)$ أو $(\propto = 0.01)$ أو$(\propto = 0.10)$.

2- نضع على متوسط المجتمع μ (أو النسبة R فيه) فرضيتين متنافيتين ومتكاملتين كما يلي:

أ- فرضية العدم: نفترض أن متوسط المجتمع μ يساوي قيمة معلومة μ_0، وهذا يعني أنه لا يوجد فرق معنوي بينه وبين القيمة المفترضة μ_0 (أي عدم وجود فرق بينهما) ونكتب ذلك كما يلي:

$$H_0: \mu = \mu_0 \qquad (9-6)$$

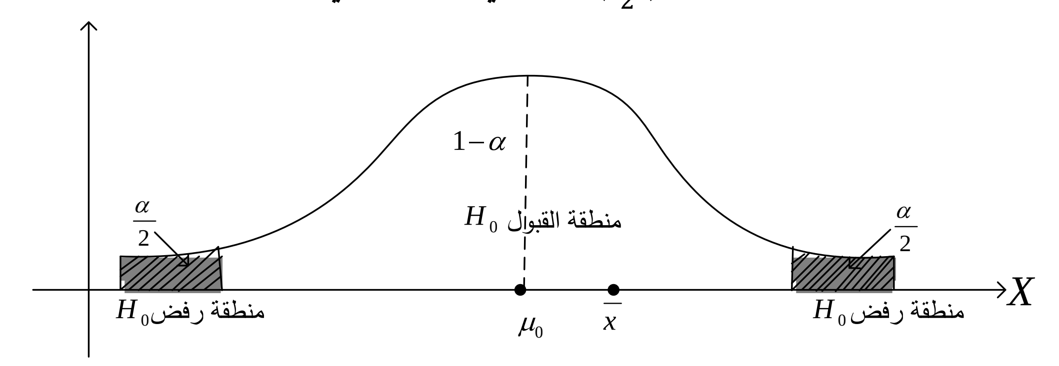
وتكون هذه الفرضية مقبولة إذا كان الفرق $(\mu - \mu_0)$ أو تقديره $(\bar{x} - \mu_0)$ واقعاً ضمن مجال الثقة المحدد للفرق $(\bar{x} - \mu)$, ونقرر ذلك من خلال مؤشر الاختبار المناسب.

ب- الفرضية البديلة: وهي الفرضية المعاكسة لفرضية العدم، ومن شكلها تتحدد منطقة الرفض, ويمكن أن تُكتب على أحد الأشكال الثلاثة التالية:

- الشكل الأول: الشكل الثنائي أو شكل عدم التساوي وتكتب الفرضية البديلة فيه كما يلي:

$$H_1: \mu \neq \mu_0 \qquad (10-6)$$

وفيه تكون منطقة الرفض واقعة على الجانبين، ولذلك يسمى هذا الشكل بالاختبار ثنائي الجانب، لأنه يخصص لكل جانب نصف مستوى الدلالة($\frac{\propto}{2}$) , كما في الشكل التالي :



الشكل (6-4): منطقة القبول ومنطقتي الرفض على اليمين واليسار

- الشكل الثاني: الأحادي اليميني، وتكتب الفرضية البديلة H_1 كما يلي:

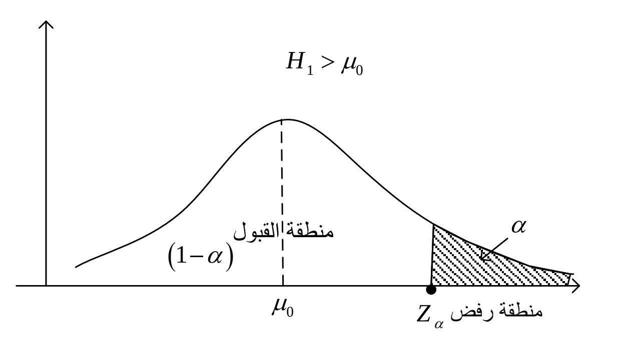
$$H_1: \mu > \mu_0 \qquad (11-6)$$

وفيه تكون منطقة الرفض على اليمين فقط، وتقابل كامل الاحتمال $\propto$ كما على الشكل (6-5).

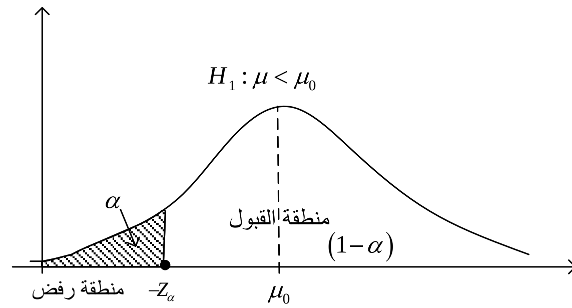
- الشكل الثالث: الأحادي اليساري، وتكتب الفرضية البديلة H_1 كما يلي:

$$H_1: \mu < \mu_0 \qquad (12-6)$$

- وفيه تكون منطقة الرفض على اليسار فقط، وتقابل كامل الاحتمال $\propto$ كما على الشكل (6-6).



الشكل (6-5): أحادي يميني



الشكل (6-6): أحادي يساري

وللتحقق من صحة أو عدم صحة فرضية العدم H_0، يجب علينا أن نسحب عينة عشوائية من المجتمع ونحسب متوسطها $\bar{x}$ ثم نقارنه مع متوسط المجتمع المفترض في الفرضية H_0 وهو μ_0. فإذا كان $\bar{x}$ يساويه أو قريباً منه نقبل فرضية العدم H_0، ونعترف بأن متوسط المجتمع يساوي μ_0، أما إذا كان $\bar{x}$ بعيداً عن μ_0 (يوجد فرق جوهري بينهما)، فإننا نرفض فرضية العدم H_0 ونقبل الفرضية H_1، ونعترف بأن متوسط المجتمع μ لايساوي μ_0، بل يساوي قيمة أخرى أكبر أو أصغر منها. وحتى لا تكون الأمور مزاجية فإن عملية مقارنة $\bar{x}$ مع μ_0، تحتاج إلى أداة إحصائية ورياضية تحدد لنا مقدار الفرق المقبول ومقدار الفرق المعنوي أو الجوهري، وهذه الأداة تسمى مؤشر الاختبار. وهكذا نجد أنفسنا بحاجة قبل كل شيء إلى سحب عينة عشوائية من المجتمع المدروس وحساب مؤشراتها المختلفة .

3- نقوم بتحديد طريقة سحب العينة (مع الإعادة أم بدون إعادة) , وحسب طريقة السحب المختارة نقوم بحساب حجم العينة من إحدى العلاقتين الخاصتين بتقدير المتوسط (أو النسبة R ضمن القوسين) وهما: [انظر الفقرة (4-4) في الفصل الرابع] .

$$n = \frac{Z^2 s^2}{d^2} = \left(\frac{Z^2 * r(1-r)}{d^2}\right) \qquad \text{للسحب مع الإعادة} \qquad (13-6)$$

$$n = \frac{NZ^2 s^2}{Nd^2 + Z^2 s^2} = \left(\frac{NZ^2 r(1-r)}{Nd^2 + Z^2 r(1-r)}\right) \qquad \text{للسحب بدون اعادة} \qquad (14-6)$$

حيث أن : s^2 هو تباين العينة أو تقديره من أي عينة سابقة .

وأن: d هو مقدار الدقة المطلوبة وتحدد من قبل الجهات المعينة أومن قبل الباحث .

وأن: Z هي قيمة المتحول الطبيعي المعياري المقابل لنصف مستوى الدلالة $\frac{\alpha}{2}$ على الطرفين.

وأن: r هو مقدار النسبة في المجتمع أو اي تقدير لها من خلال أي عينة اختبارية.

وفي حالة اختبار النسب المتوازنة في المجتمع نضع ($r = 0.50$) حتى نحصل على أكبر حجم ممكن للعينة، أما عندما يكون حجم المجتمع N كبيراً أو غير معروف، يفضل استخدام العلاقة (6-13) للسحب مع الإعادة, ثم نقوم بسحب العينة المذكورة من المجتمع ونحسب متوسطها $\bar{x}$ وتباينها s^2 من العلاقتين:

$$\bar{x} = \frac{1}{n}\sum x_i \qquad (15-6)$$

$$s^2 = \frac{1}{n-1} * \sum (x_i - \bar{x})^2 \qquad (16-6)$$

4- نقوم بحساب مؤشر الاختبار للمتوسط (أو للنسبة ضمن القوسين) من العلاقة المعيارية التالية:

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \left(\frac{|r - R_0|}{\sqrt{\frac{R_0(1 - R_0)}{n}}} \right) \quad : \quad (\sigma \text{ معلوم}) \qquad (17-6)$$

وهو متحول عشوائي يخضع للتوزيع الطبيعي المعياري $N(0,1)$، ولكن بما أن تباين المجتمع σ^2 وبالتالي انحرافه المعياري σ يكون غالباً مجهولاً، فإن حساب قيمة Z السابقة يكون أمراً مستحيلاً. ولكي نتخلص من هذه المشكلة نستبدل σ^2 بتباين العينة s^2 كتقدير جيد له، ونعرف مؤشر جديد لاختبار متوسط المجتمع μ (دون تعديل المقام في مؤشر النسبة لأن R_0 تكون معلومة من فرضية العدم H_0). بالعلاقة التالية :

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \left(\frac{|r - R_0| - \frac{1}{2n}}{\sqrt{\frac{R_0(1 - R_0)}{n}}} \right) \qquad (18-6)$$

وهو متحول جديد يخضع لتوزيع (ستودينت) بـ $(n-1)$ درجة حرية عند اختبار المتوسط, وللتوزيع الطبيعي عند اختبار النسبة R .

ملاحظة: إذا كان حجم العينة n كبيراً $(n > 30)$، فإن توزيع (ستودينت) يقترب من التوزيع الطبيعي المعياري، وعندها نعتبر t في العلاقة (6-18) خاضعاً للتوزيع الطبيعي المعياري.

5- نقوم بتحديد منطقتي الرفض والقبول واتخاذ القرار:

لتحديد منطقتي الرفض والقبول ولاتخاذ القرار اللازم حول صحة الفرضية H_0، يوجد طريقتان لاتخاذ القرار المناسب هما: طريقة القيمة الحرجة، وطريقة احتمال الدلالة P . وسنشرحهما كما يلي:

أ- **طريقة القيمة الحرجة** لـ Z أو t : لنفترض أن الاختبار ثنائي الجانب (أي أن $H_1: \mu \neq \mu_0$)، فعندها يكون مستوى الدلالة $\propto$ موزعاً على الجانبين، وهنا يكون لدينا حالتان لـ σ هما: إما أن تكون قيمة σ معلومة، أو أن تكون σ مقدرة من العينة بـ s ، ولذلك نعالجها كما يلي:

- إذا كانت قيمة σ معلومة: فإننا نحسب قيمة مؤشر الاختبار من العلاقة (6-17) ثم نقارنها مع القيمة الحرجة لمتحول التوزيع الطبيعي $Z_{\frac{\propto}{2}}$، لذلك نبحث في جدول التوزيع الطبيعي المعياري عن القيمة الحرجة $Z_{\frac{\propto}{2}}$ المقابلة لنصف مستوى الدلالة $\frac{\propto}{2}$ على الطرفين، ثم نقارن القيمة المحسوبة Z معها، ونتخذ القرار حول H_0 كما يلي:

- إذا كانت $|Z| \leq Z_{\frac{\propto}{2}}$، تكون Z واقعة في منطقة القبول, وعندها نقبل فرضية العدم ونقول بأن $\mu = \mu_0$ ، باحتمال ثقة يساوي $1-\propto$.
- إذا كانت $|Z| > Z_{\frac{\propto}{2}}$، تكون Z واقعة في منطقة الرفض، وعندها نرفض فرضية العدم ونقول بأن $\mu \neq \mu_0$ ، بمستوى دلالة يساوي $\propto$.

- أما عندما يكون σ^2 مجهولاً, فإننا نستخدم تقديره s ونحسب قيمة المؤشر t من العلاقة (6−18) , ثم نبحث في جدول توزيع (ستودينت) عن القيمة الحرجة $t_{\frac{\propto}{2}}$ المقابلة لنصف مستوى الدلالة $\frac{\propto}{2}$ على الطرفين ولدرجة حرية $(n-1)$ ونتخذ القرار بالمقارنة كما يلي:
  - إذا كانت $|t| \leq t_{\frac{\propto}{2}}$ نقبل فرضية العدم ونقول بأن $\mu = \mu_0$ باحتمال ثقة $1-\propto$.
  - إذا كانت $|t| > t_{\frac{\propto}{2}}$ نرفض فرضية العدم ونقول بأن $\mu \neq \mu_0$ بمستوى دلالة $\propto$.

ملاحظة: إذا كان الاختبار أحادي الجانب (يميني ويساري) , فإن القيمة الحرجة لـ Z تكون هي القيمة المقابلة لكامل مستوى الدلالة $\propto$ ونرمز لها بـ $Z_{\propto}$، وعندها نتخذ القرار عند المقارنة حول H_0 كما يلي:

- إذا كان الاختبار أحادي يميني $(H_1: \mu > \mu_0)$ فإننا نقارن قيمة Z المحسوبة مع $Z_{\propto}$, فإذا كانت $Z \leq Z_{\propto}$ نقبل فرضية العدم H_0، ونقبل بأن $\mu = \mu_0$، أما إذا كانت $Z > Z_{\propto}$ فإننا نرفض H_0 ونعترف بأن $\mu > \mu_0$ كما في الشكل (6−5) السابق.
- أما إذا كان الاختبار أحادي يساري $(H_1: \mu < \mu_0)$، فإننا نقارن قيمة Z المحسوبة مع $(-Z_{\propto})$ السالبة، فإذا كانت $(Z \geq -Z_{\propto})$ نقبل فرضية العدم H_0 ونقبل بأن $\mu = \mu_0$، أما إذا كانت $(Z < -Z_{\propto})$، فإننا نرفض H_0 ونقبل H_1 ونعترف بأن $\mu < \mu_0$ كما في الشكل (6−6) السابق.

وكذلك الأمر عند استخدامنا لمؤشر (ستودينت) t فإن القيمة الحرجة لمتحوله t هي القيمة المقابلة لكامل مستوى الدلالة $\propto$ ولدرجة الحرية $(n-1)$ ونرمز لها بـ $t_{\propto}$. ونتخذ القرار كما يلي:

- إذا كان الاختبار أحادي يميني $(H_1: \mu > \mu_0)$ وكان $t \leq t_{\propto}$ فإننا نقبل فرضية العدم H_0 ونقبل بأن $\mu = \mu_0$، أما إذا كان $t > t_{\propto}$ فإننا نرفض H_0 ونقبل H_1 ونعترف بأن $\mu > \mu_0$.
- أما إذا كان الاختبار أحادي يساري $(H_1: \mu < \mu_0)$ فإننا نقارن قيمة t المحسوبة مع $(-t_{\propto})$ السالبة، فإذا كانت $(t \geq -t_{\propto})$ فإننا نقبل الفرضية H_0 ونقبل بأن $\mu = \mu_0$، أما إذا كانت $(t < -t_{\propto})$ فإننا نرفض H_0 ونقبل H_1 التي تقول بأن: $\mu < \mu_0$.

مثال (6−1): لنفترض إننا نريد اختبار أن يكون توقع X في المجتمع $\mu_0 = 50$، فسحبنا عينة عشوائية منه بحجم $n = 25$ عنصراً . فكان متوسطها: $\bar{x} = 53$ وتباينها: $s^2 = 400$, ثم حددنا مستوى الدلالة بـ $\propto = 0.05$ ووضعنا الفرضيتين كما يلي:

$H_0: \mu = 50 \qquad H_1: \mu \neq 50$ (الاختبار ثنائي الجانب)

ولإجراء هذا الاختبار نلاحظ أن تباين المجتمع σ^2 مجهول, لذلك نستخدم تقديره s ونحسب قيمة t من العلاقة (6−18) والتي تأخذ الشكل التالي :

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{53 - 50}{20/5} = \frac{3}{4} = 0.75$$

وبما أن t يخضع لتوزيع (ستودينت) بـ $(n-1)$ درجة حرية , فإننا نقوم بحساب القيمة الحرجة له $t_{n-1}\left(\frac{\propto}{2}\right)$ ، فنجد من جداول (ستودينت) أن:

$$t_{n-1}\left(\frac{\propto}{2}\right) = t_{24}\left(\frac{0.05}{2}\right) = t_{24}(0.025) = 2.064$$

وبمقارنة القيمة المحسوبة $t = 0.75$ مع القيمة الحرجة $t_{24}\left(\frac{\propto}{2}\right) = 2.064$ نجد أن $|t| < t_{24}\left(\frac{\propto}{2}\right)$ لذلك نقبل فرضية العدم H_0 والتي تقول أن $\mu = 50$, وذلك باحتمال ثقة $1-\propto = 0.95$.

ملاحظة: إذا كان تباين المجتمع σ^2 معلوماً نطبق العلاقة (6−7) ونستخدم المتحول المعياري Z .

ب− **طريقة احتمال الدلالة P** (P- Value) أو (Sig):

لتوضيح كيفية تطبيق هذه الطريقة لابد من توضيح معنى احتمال الدلالة (significance probability) والذي يرمز له في البرامج الحاسوبية بالرمز P أو(P-value) أو بالرمز (Sig). إن احتمال الدلالة P حسب التعريف هو: الاحتمال الذي تتركه القيمة **المحسوبة** لمؤشر الاختبار Z أو t أو غيرهما على طرفي التوزيع الاحتمالي، أو على أحد طرفيه، وهو يتأثر بنوع الاختبار وبحالاته المختلفة التالية:

فإذا كان الاختبار ثنائي الجانب $(H_1: \mu \neq \mu_0)$، فإن قيمة الاحتمال P، يتم توزيعها بالتساوي على طرفي التوزيع، بحيث يكون لكل طرف $\frac{P}{2}$، والطرفان هنا يقابلان في التوزيع الطبيعي المعياري المجالين المفتوحين $]+Z , +\infty[$ و $]-\infty , -Z[$.

أما إذا كان الاختبار أحادي يميني $(H_1: \mu > \mu_0)$، فإن كامل قيمة P تكون متوضعة على اليمين وتقابل المجال المفتوح $]+Z , +\infty[$.

وإذا كان الاختبار أحادي يساري $(H_1 : \mu < \mu_0)$، فإن كان كامل قيمة P تكون متوضعة على اليسار وتقابل المجال المفتوح $]-\infty , -Z[$.

أي أن الاحتمال P يساوي المساحة التي تقع تحت منحني التوزيع وتقابل المجالات المذكورة (حسب كل حالة)، ويتم تحديد القيمة العددية لـ P بعد إعداد الحسابات اللازمة والقيام بإجراء الاختبار المفروض والحصول على القيمة المحسوبة Z أو t أو غيرهما . ثم حساب قيمة تكامل التوزيع الاحتمالي على المجالات المذكورة (حسب كل حالة)، كما سنرى لاحقاً.

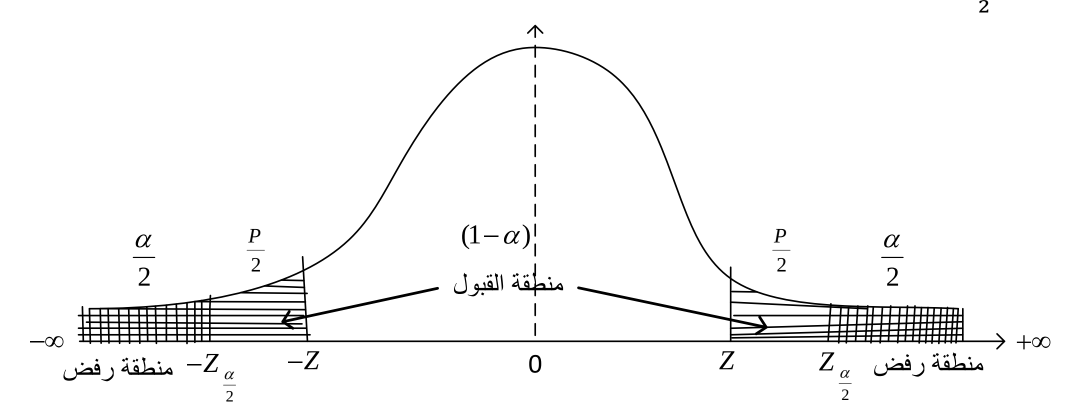
واخيراً نشير إلى أن احتمال الدلالة P يختلف جذرياً عن مستوى الدلالة ∝ (Level of significance), الذي يحدده الباحث أو المشرفون على البحث قبل إجراء البحث وقبل إجراء الاختبار نفسه.

ويستفاد من الاحتمال P في اتخاذ القرارات حول الفرضية H_0 وذلك بمقارنتها مع مستوى الدلالة ∝.

ولتوضيح ذلك نأخذ حالة التوزيع الطبيعي المعياري، ونحسب منه قيمة احتمال الدلالة P ، حسب حالات الاختبار التالية:

1) حالة الاختبار ثنائي الجانب: أي أن منطقة الرفض حسب القواعد السابقة تقع على الجانبين.

– فإذا كانت $|Z| \leq Z_{\frac{\propto}{2}}$ فإننا نقبل فرضية العدم H_0 ويكون لدينا الشكل التالي:



شكل (6–7): تحديد المساحة P

من هذا الشكل نلاحظ أن $\frac{\propto}{2}$ هي المساحة المظللة بخطوط عمودية وهي التي تقابل المجال$\left]Z_{\frac{\propto}{2}}, +\infty\right[$.

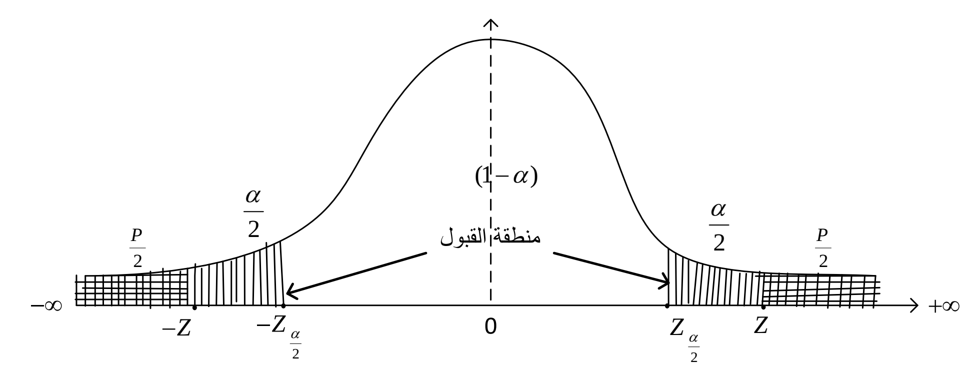
أما $\frac{P}{2}$ فهي المساحة المظللة بخطوط أفقية وهي التي تقابل المجال $]Z , +\infty[$، وعندما تكون $Z < Z_{\frac{\propto}{2}}$ كما في الشكل (6–7) يكون لدينا $\frac{P}{2} > \frac{\propto}{2}$ ، ويكون لدينا $P >\propto$، أي أنه علينا أن نقبل الفرضية H_0 إذا كانت $P >\propto$ (لأنه يكون لدينا $|Z| < Z_{\frac{\propto}{2}}$) .

ويتم حساب قيمة الاحتمال P في هذه الحالة من تكامل التوزيع الطبيعي المعياري على المجالين المتناظرين $]Z , +\infty[$ و$]-\infty , -Z[$ كما يلي:

$$P = \int_{-\infty}^{-Z} f(x)dx + \int_{Z}^{+\infty} f(x)dx = 2 * \int_{Z}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} * dx = 2[1 - \phi(Z)] \quad (19-6)$$

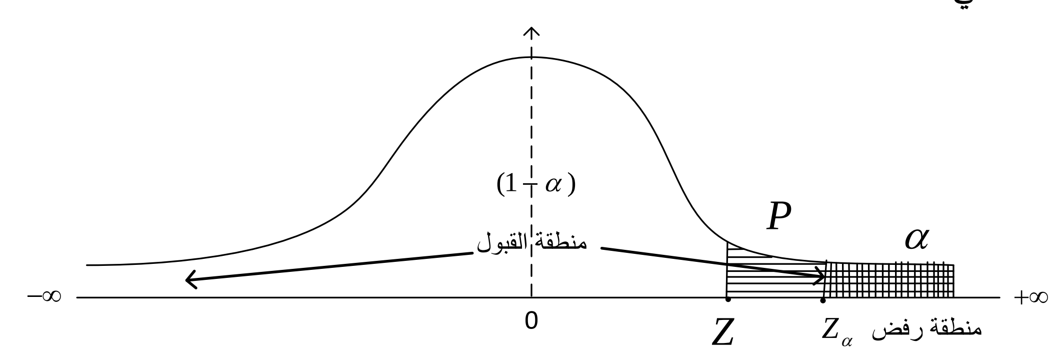
ويمكن الحصول على قيمة هذا التكامل من الجداول الإحصائية الجاهزة أو من الحواسيب المبرمجة .

– أما إذا كانت $|Z| > Z_{\frac{\propto}{2}}$ فإننا نرفض فرضية العدم H_0، وعندها يكون لدينا $\frac{P}{2} < \frac{\propto}{2}$ ، أي يكون لدينا $P <\propto$، أي أنه يجب علينا أن نرفض فرضية العدم H_0 إذا كان $P <\propto$ (لأنه تكون $|Z| > Z_{\frac{\propto}{2}}$) كما هو موضح في الشكل (6–8) التالي، حيث رمزنا للمساحة المظللة بخطوط أفقية من الطرفين للاحتمال P وللمساحة المظللة بخطوط عمودية لمستوى الدلالة ∝.



شكل (6−8): تحديد المساحة P

2) حالة الاختبار الأحادي اليميني: فإذا كانت $Z \leq Z_{\propto}$ فإننا نقبل فرضية العدم H_0، ويكون لدينا الشكل التالي:



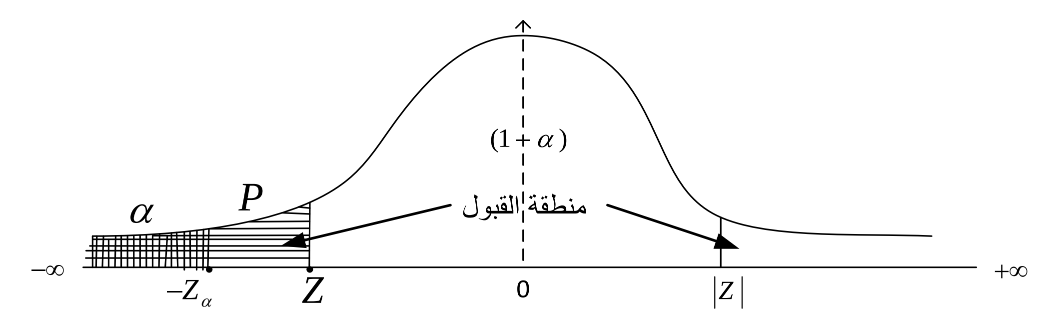
الشكل (6−9) تحديد المساحة P

ومن الشكل (الشكل (6−9) نلاحظ أن ∝ هي المساحة المظللة بخطوط عمودية وهي التي تقابل المجال$]Z_{\propto}, +\infty[$، أما P فهي المساحة المظللة بخطوط أفقية وهي التي تقابل المجال $]Z, +\infty[$، وهنا يكون لدينا $P >\propto$، أي أنه علينا أن نقبل فرضية العدم H_0 إذا كان $P >\propto$ (لأن $Z \leq Z_{\propto}$).

− أما إذا كانت $Z > Z_{\propto}$ (تقع على يمينها) فإننا نرفض الفرضية H_0، وعندها يكون لدينا $P <\propto$، أي أنه علينا أن نرفض الفرضية H_0 إذا كان $P <\propto$ (لأن $Z > Z_{\propto}$)، ويتم حساب قيمة P في هذه الحالة من تكامل التوزيع الطبيعي المعياري على المجال$]Z, +\infty[$ كما يلي :

$$P = \int_{Z}^{+\infty} f(x) d_x = \int_{Z}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} * dx = 1 - \phi(Z) \qquad (20-6)$$

3) حالة الاختبار الأحادي اليساري: إذا كانت $Z > -Z_{\propto}$ فإننا نقبل فرضية العدم H_0 ويكون لدينا الشكل التالي: (مع ملاحظة أن قيمة Z هي قيمة جبرية فقد تكون سالبة أو موجبة)



الشكل (6− 10): تحديد المساحة P

ومن الشكل (6− 10) نلاحظ أن $\propto$ هي المساحة المظللة بخطوط عمودية وهي التي تقابل المجال $]-\infty,-Z_{\propto}[$، أما قيمة الاحتمال P فهو المساحة المظللة بخطوط أفقية وهي المساحة التي تقابل المجال $]-\infty\ ,\ Z[$، وهذا يعني أن $P >\propto$، لذلك يجب علينا أن نقبل الفرضية H_0 إذا كان $P >\propto$ (لأن $Z > -Z_{\propto}$) . [انتبه إلى ذلك الاختلاف].

− أما إذا كانت $Z < -Z_{\propto}$ (تقع على يسارها) فإننا نرفض الفرضية H_0، وعندها يكون لدينا $P <\propto$، أي أنه علينا أن نرفض الفرضية H_0 إذا كانت $P <\propto$ (لأن $Z < -Z_{\propto}$)، ويتم حساب P في هذه الحالة من العلاقة التالية:

$$P = \int_{-\infty}^{Z} f(x)d_x = \int_{|Z|}^{+\infty} f(x)d_x = \int_{|Z|}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} * dx = \phi(Z) \quad (21-6)$$

ملاحظة: إذا كان مؤشر الاختبار يخضع لتوزيع (ستودينت) t أو لأي توزيع آخر مثل X^2 أو F، فإن قيمة P تحسب حاسوبياً من تكاملات مشابهة للتكاملات السابقة على تلك التوزيعات وعلى المجالات المناسبة والمشابهة لتلك المجالات المذكورة. وهي أمور كثيرة وطويلة لا مجال للخوض فيها في هذا الفصل.

نتيجة هامة: نلاحظ مما سبق أنه يمكننا استخلاص النتيجة التالية:

إذا كانت $P >\propto$ فإننا نقبل فرضية العدم H_0 مهما كان شكل الاختبار(ثنائي الجانب أم أحادي يميني أو يساري)، أما إذا كانت $P <\propto$ فإننا نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1 بمستوى دلالة أقل من $\propto$(ويساوي P). وهذا ماجعل طريقة الاحتمال P أكثر استخداماً في البرامج الحاسوبية رغم محاذيرها، التي تجعل الباحث يرفض الوجبة(H_0) من رائحتها دون أن يتذوقها(يفهمها).

مثال (6−2) : لنفترض إننا نريد التأكد من نتيجة الاختبار في المثال (6−1), لذلك قمنا بسحب عينة أخرى كبيرة بحجم $n = 100$ عنصراً، ثم حسبنا متوسطها وتباينها فكانا يساويان ما يلي: $\bar{x} = 54$ و$s^2 = 225$, ثم حددنا مستوى الدلالة بـ ($\propto= 0.05$) ووضعنا الفرضيتين كما يلي:

$H_0: \mu = 50 \qquad H_1: \mu \neq 50$ (الاختبار ثنائي الجانب)

وهنا نلاحظ أيضاً أن تباين المجتمع σ^2 مجهول . لذلك يجب علينا أن نطبق العلاقة (6−18)، ولكن بما أن حجم العينة n كبيراً $(n > 30)$ فإن تلك العلاقة تقترب من العلاقة (6−17) ويصبح t متقارباً مع Z ونكتبها كما يلي:

$$t \approx Z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{54 - 50}{15/5} = \frac{4}{1.5} = 2.667$$

وبما أن Z يخضع للتوزيع الطبيعي المعياري، فإننا نقوم بإيجاد القيمة الحرجة $Z\left(\frac{\propto}{2}\right)$ من جداول التوزيع الطبيعي المعياري فنجد أن:

$$Z\left(\frac{\propto}{2}\right) = Z\left(\frac{0.05}{2}\right) = Z(0.025) = 1.96$$

وبمقارنة قيمة $Z = 2.667$ المحسوبة مع $Z\left(\frac{\propto}{2}\right) = 1.96$ الحرجة نجد أن:

$|Z| > Z\left(\frac{\propto}{2}\right)$. لذلك نرفض فرضية العدم H_0 التي تقول أن $\mu = 50$ ونقبل الفرضية H_1 التي تقول أن توقع X في ذلك المجتمع يختلف عن 50 ، وذلك باحتمال ثقة 0.95 .

ملاحظة:كان يمكننا أن نستخدم طريقة احتمال الدلالة P لاتخاذ القرار حول H_0 ، ولذلك يجب أن نقوم بعد حساب Z ، بحساب الاحتمال P من العلاقة (6−19)، فنجد من جداول التوزيع الطبيعي المعياري أن:

$$P = 2[1 - \phi(Z)] = 2[1 - \phi(2.667)] = 2[1 - 0.99615] = 0.0077$$

وبمقارنة P المحسوبة(0.0077) مع $\propto$ المفروضة بـ(0.05) نجد أن $P < \propto$. لذلك نرفض فرضية العدم H_0 بمستوى دلالة أقل من $\propto$ (ويساوي0.0077)، أي باحتمال ثقة أكبر من 0.95(ويساوي0.9923).

وهنا نلاحظ أن نتيجة الاختبار في هذا المثال تختلف عن نتيجة المثال (6−1) وذلك لأن العينة الأخيرة مختلفة عن العينة الأولى ببياناتها وبحجمها.

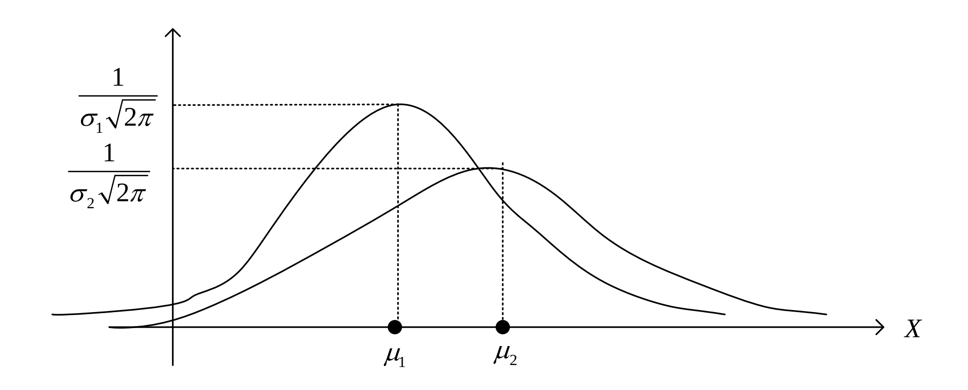
6−4: اختبارات معالم مجتمعين طبيعيين (من عينتين مستقلتين):

6−4−1: اختبار الفرق بين متوسطي مجتمعين طبيعيين:

لنفترض أننا نريد دراسة تغيرات متحول طبيعي X في مجتمعين منفصلين:

ولنفترض أن توقع X في المجتمع الأول هو μ_1 وتباينه فيه σ_1^2 وإن توزيعه الطبيعي هو $N(\mu_1, \sigma_1^2)$، كما نفترض أن توقع X نفسه في المجتمع الثاني هو μ_2 وتباينه فيه σ_2^2 وإن توزيعه الطبيعي هو $N(\mu_2, \sigma_2^2)$.

وبرسم هذين التوزيعين على شكل واحد نحصل على الشكل التالي:



الشكل (6-11): شكلان طبيعيان فيهما $\sigma_1^2 < \sigma_2^2$

ومن هذا الشكل نلاحظ أن عملية المقارنة بين μ_1 و μ_2 لا تتعلق بالفرق بينهما $(\mu_1 - \mu_2)$ فقط . بل تتعلق بشكل التوزيع الطبيعي وبتبايني X في هذين التوزيعين.

فإذا كان σ_1^2 و σ_2^2 مختلفين كثيراً فإن عملية المقارنة لا تكون متوازنة، لأنه إذا كان $\sigma_1^2 < \sigma_2^2$ فإن قيم X تكون متمركزة حول μ_1 أكثر من تمركز قيم X حول μ_2 .

وإن عملية المقارنة بين التوقعين μ_1 و μ_2 تكون أكثر فعالية عندما يكون $\sigma_1^2 = \sigma_2^2$ (أي عندما يكون الشكلان متشابهين) لذلك فإننا سنميز بين الحالتين التاليتين:

الحالة الأولى: الحالة التي يكون فيها $\sigma_1^2 \neq \sigma_2^2$

الحالة الثانية: الحالة التي يكون فيها $\sigma_1^2 = \sigma_2^2$

كما أننا سندرس الحالة التي يكون فيها σ_1^2 و σ_2^2 معلومين والحالة التي يكون فيها σ_1^2 و σ_2^2 مجهولين.

ولإجراء هذا الاختبار حول $(\mu_1 - \mu_2)$ علينا أن نسحب من هذين المجتمعين عينيتين عشوائيتين ومستقلتين بحجمين n_1 و n_2 ونحسب منهما ما يلي:

أ- متوسط العينة الأولى: $\bar{x}_1 = \frac{\sum x_i}{n_1}$, ويعتبر تقديراً غير متحيز لمتوسط المجتمع الأول μ_1 .

ب- متوسط العينة الثانية: $\bar{x}_2 = \frac{\sum x_i}{n_2}$, ويعتبر تقديراً غير متحيز لمتوسط المجتمع الثاني μ_2 .

ج- ثم نحسب التباين المصحح للعينة الأولى من العلاقة: $s_1^2 = \frac{1}{n_1 - 1}\sum(x_{i1} - \bar{x}_1)^2$,
ويعتبر هذا التباين تقديراً غير متحيز لتباين المجتمع الأول σ_1^2 .

د- ونحسب التباين المصحح للعينة الثانية من العلاقة : $s_2^2 = \frac{1}{n_2 - 1}\sum(x_{i2} - \bar{x}_2)^2$، ويعتبر هذا التباين تقديراً غير متحيز لتباين المجتمع الثاني σ_2^2 .

هـ-ثم نحسب الفرق بين متوسطي هاتين العينتين $(\bar{x}_1 - \bar{x}_2)$، وهو يعتبر تقديراً غير متحيز للفرق بين متوسطي المجتمعين$(\mu_1 - \mu_2)$, كما يعتبر الفرق $(\bar{x}_1 - \bar{x}_2)$ متحولاً عشوائياً جديداً يخضع للتوزيع الطبيعي الذي توقعه $(\mu_1 - \mu_2)$ وتباينه $Var(\bar{x}_1 - \bar{x}_2)$.

ولاختبار الفرق بين متوسطي المجتمعين $(\mu_1 - \mu_2)$ نضع الفرضيتين كما يلي:

فرضية العدم: $H_0: (\mu_1 - \mu_2) = 0$

وتقابلها الفرضية البديلة والتي يمكن أن تكون على أحد الأشكال التالية:

على الشكل الثنائي الجانب : $H_1: (\mu_1 - \mu_2) \neq 0$.

أو على الشكل الأحادي اليميني: $H_1: (\mu_1 - \mu_2) > 0$.

أو على الشكل الأحادي اليساري: $H_1: (\mu_1 - \mu_2) < 0$.

ثم نقوم بتشكيل مؤشر الاختبار المعياري لمتحول الفرق $(\bar{x}_1 - \bar{x}_2)$ من العلاقة التالية :

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{Var(\bar{x}_1 - \bar{x}_2)}} \qquad (19-6)$$

ثم نقوم بمعالجته حسب الحالات السابقة لـ σ_1^2 و σ_2^2 التالية:

الحالة الأولى: إذا كان $\sigma_1^2 \neq \sigma_2^2$ وكان σ_1^2 و σ_2^2 معلومين عددياً فإننا نجد أن تباين الفرق $(\bar{x}_1 - \bar{x}_2)$ لهاتين العينتين المستقلتين يساوي [لعدم وجود ارتباط بين العينتين] ما يلي:

$$Var(\bar{x}_1 - \bar{x}_2) = Var(\bar{x}_1) + Var(\bar{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \ (\text{معلوم})$$

وبالتعويض في (6–19) نحصل على مؤشر الاختبار الطبيعي المعياري التالي :

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad : \left(\sigma_1^2 \neq \sigma_2^2 \ \text{ومعلومين}\right) \qquad (20-6)$$

وهو متحول يخضع للتوزيع الطبيعي المعياري $N(0\,,1)$.

ولإجراء الاختبار نحسب قيمة المؤشر Z من العلاقة (6–20)، وإذا كان الاختبار ثنائي الجانب نقارنها مع قيمة $Z_{\frac{\alpha}{2}}$ الحرجة والمقابلة لنصف مستوى الدلالة $\frac{\alpha}{2}$ من الطرفين ، وبناء على هذه المقارنة نقبل أو نرفض فرضية العدم H_0 وفق القواعد المذكورة سابقاً.

أما عندما يكون التباينان σ_1^2 و σ_2^2 مجهولين وغير متساويين، فإننا نستبدلهما بتقديريهما غير المتحيزين s_1^2 و s_2^2 والمحسوبين من العينتين فنحصل على مؤشر اختبار آخر t يخضع تقاربياً لتوزيع (ستودينت) وله درجة حرية معقدة (انظر المثال (6–3)) وهو يأخذ الشكل التالي:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad : \left(\sigma_1^2 \neq \sigma_2^2 \ \text{ومجهولين}\right) \qquad (21-6)$$

ملاحظة: يمكن استخدام مؤشر الاختبار الأخير t في اختبارات الفرق بين المتوسطين $(\mu_1 - \mu_2)$ إذا كان حجما العينتين n_1 و n_2 كبيرين، وعندها نعتبر درجة الحرية مساوية لأصغر العددين $(n_1 - 1)$ أو $(n_2 - 1)$ لأن قيمتها الحقيقية تكون قريبة منها.

الحالة الثانية: وهي الحالة التي يكون فيها تباينا المجتمعين متساويين ومعلومين, أي عندما يكون لدينا $\sigma_1^2 = \sigma_2^2 = \sigma^2$, حيث σ^2 هي القيمة الموحدة والمعلومة لهما، وعندها تأخذ العلاقة (6− 20) الشكل التالي:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sigma * \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad : \left(\sigma^2 \text{معلوم}\right) \qquad \mathbf{(22-6)}$$

وهو متحول يخضع للتوزيع الطبيعي المعياري $N(0\,,1)$, ونتعامل معه كما في العلاقة (6−20).

أما عندما يكون التباين الموحد σ^2 مجهولاً , فإننا نقدره من خلال المتوسط الحسابي للتباينين s_1^2 و s_2^2 المصححين والمحسوبين من العينتين والمثقلين بـ $(n_1 - 1)$ و $(n_2 - 1)$ على الترتيب، فنحصل من العلاقة المركبة لهما على ما يسمى بالتباين المدمج pooled ونرمز له بالرمز S_P^2 ونكتبه كما يلي:

$$S_P^2 = \widetilde{\sigma^2} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \qquad (23-6)$$

ويبرهن في الاحصاء الرياضي على أن التقدير S_P^2 هو تقدير غير متحيز للتباين الموحد σ^2، وبذلك تأخذ العلاقة (6−22) الشكل التالي:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad (24-6)$$

أو الشكل التالي:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{S_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad (25-6)$$

حيث استبدلنا في (6−22) التباين المشترك σ^2 بالتباين المدمج S_P^2 ، ومنها نحصل على المؤشر t الخاضع تقاربياً لتوزيع (ستودينت) بـ $(n_1 + n_2 - 2)$ درجة حرية، ويمكن استخدامه في اختبارات الفروق بشرط أن يكون تباينا المجتمعين متساويين، لذلك يجب أن نتأكد أولاً من تحقق الشرط السابق: $\sigma_1^2 = \sigma_2^2 = \sigma^2$ قبل تطبيق (6−25)، و ذلك بإجراء إختبار تساوي التباينين كما سنرى لاحقاً.

ولإجراء الاختبار نحسب قيمة المؤشر t من العلاقة (6−25)، ثم نقارنها مع قيمة $t_{\frac{\propto}{2}}$ الحرجة والمقابلة لنصف مستوى الدلالة $\frac{\propto}{2}$ من الطرفين ولدرجة الحرية $(n_1 + n_2 - 2)$، وبناء على هذه المقارنة نقبل أو نرفض فرضية العدم H_0 وفق القواعد المذكورة سابقاً.

ولكنه عندما يكون $(n_1 + n_2 - 2) > 30$ كبيراً, فإنه يخضع تقاربياً للتوزيع الطبيعي المعياري, ونتعامل معه كما في العلاقة (6−20).

ملاحظة :

إذا كانت نتيجة الاختبار رفض فرضية العدم H_0 لعدم تساوي بعض المتوسطات، فإنه لا بد لنا من أن نتعرف أو نحدد مصادر عدم التساوي , لذلك نقوم بإجراء مقارنات ثنائية بين كل متوسطين من مجتمعين مختلفين ونطبق أحد الاختبارين التاليين:

a. **اختبار بونفيروني Bonferroni's Test** الذي يعتمد على المؤشر التالي:

$$t = \frac{(x_i - \overline{x_j}) - 0}{MSSE \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} \qquad (25a\text{-}6)$$

$i \neq j$, $i , j = 1 , 2 , 3 \ldots\ldots\ldots k$

حيث يتم حساب المقدار MSSE من العلاقة:

$$MSSE = \sqrt{\frac{(n_i - 1)\, s_i^2 + (n_j - 1)\, s_j^2}{n_i + n_j - 2}}$$

علماً بأن القيمة الحرجة لـ Z أو لـ t تقابل مستوى دلالة يساوي $\frac{\alpha}{m}$ ، حيث أن: m هو عدد المقارنات الممكنة ويساوي $m = c_k^2$.

b. **اختبار شفييه Schefee's Test** الذي يعتمد على المؤشر التالي:

$$t^2 = \frac{(\overline{x_i} - \overline{x_j})^2}{MSSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \qquad (25b\text{-}6)$$

وتتم مقارنة القيمة t^2 مع القيمة الحرجة $F\alpha$ مضروبة بـ $(k - 1)$.

6−4−2: اختبار الفرق بين نسبتين في مجتمعين طبيعيين:

لاختبار الفرق بين نسبتي خاصتين R_1 وR_2 في مجتمعين طبيعيين، نسحب منهما عينتين كبيرتين ، ونفترض أولاً أن $\sigma_1^2 \neq \sigma_2^2$ ونضع فرضية العدم كما يلي: $H_0: R_1 - R_2 = 0$ ، والفرضية البديلة من الشكل: $H_1: R_1 - R_2 \neq 0$ ، أو من شكل أحادي آخر. وعندها يأخذ مؤشر الاختبار الأول (6−20) الشكل التالي:

$$Z = \frac{(r_1 - r_2) - (R_1 - R_2)_0}{\sqrt{\frac{r_1(1 - r_1)}{n_1} + \frac{r_2(1 - r_2)}{n_2}}} \qquad (26 - 6)$$

حيث أن: r_1 و r_2 هما النسبتان في العينتين المسحوبتين من المجتمعين المذكورين على الترتيب.

أما عندما يكون $\sigma_1^2 = \sigma_2^2 = \sigma^2$، فإننا نقدر σ^2 من النسبة المتوسطة $\bar{r}$ ونحسب تقديره من العلاقة التالية:

$$\widetilde{\sigma^2} = \bar{r}(1 - \bar{r}) \qquad (27 - 6)$$

حيث أن النسبة المتوسطة $\bar{r}$ تحسب من المتوسط المثقل للنسبتين r_1 و r_2 كما يلي:

$$\bar{r} = \frac{n_1 r_1 + n_2 r_2}{n_1 + n_2} \qquad (28-6)$$

وعندها تأخذ العلاقة (6− 26) شكلاً آخر هو التالي:

$$t = \frac{(r_1 - r_2) - (R_1 - R_2)_0}{\sqrt{\frac{\bar{r}(1-\bar{r})}{n_1} + \frac{\bar{r}(1-\bar{r})}{n_2}}} = \frac{(r_1 - r_2) - (R_1 - R_2)_0}{\sqrt{\bar{r}(1-\bar{r})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad (29-6)$$

ومنها نحصل على المؤشر t الخاضع تقاربياً لتوزيع (ستودينت) بـ $(n_1 + n_2 - 2)$ درجة حرية,
ولكنه عندما يكون $(n_1 + n_2 - 2) > 30$، فإنه يخضع تقاربياً للتوزيع الطبيعي المعياري. ولإجراء الاختبار نحسب قيمة المؤشر t من العلاقة (6−29)، ثم نقارنها مع قيمة $t_{\frac{\alpha}{2}}$ الحرجة والمقابلة لنصف مستوى الدلالة $\frac{\alpha}{2}$ من الطرفين ولدرجة الحرية $(n_1 + n_2 - 2)$، وبناء على هذه المقارنة نقبل أو نرفض فرضية العدم H_0 وفق القواعد المذكورة سابقاً.

مثال (6−3): لدراسة حالة الفروقات بين كميتي البروتين ACTB-1 عند مرضى السرطان ومرضى الربو مقارنة مع الأشخاص الطبيعيين, أجريت التجارب اللازمة على ثلاث عينات مستقلة واستخلصت منها النتائج والبيانات التالية [من نتائج التجارب لإطروحة رسلان في إسبانيا عام 2016]:

المؤشر / العينة	حجم العينة n_i	متوسط تعبير البروتين في العينة $\bar{x}_i$	الانحراف المعياري SD_i
مرضى الربو	25	1063.126	669.1437
الأشخاص الطبيعيين	42	1535.488	479.3964
مرضى السرطان	14	2350.761	1116.602

والمطلوب: اختبار الفرق بين متوسط البروتين عند مرضى السرطان ومتوسطه عند الأشخاص الطبيعيين. ثم اختبار الفرق بين متوسطه عند مرضى الربو ومتوسطه عند الأشخاص الطبيعيين، وذلك بمستوى دلالة $\alpha = 0.05$.

الحل: لاختبار الفرق بين متوسطي البروتين ACTB-1 عند مرضى السرطان μ_1 وعند الأشخاص الطبيعيين μ_2، نضع الفرضيتين (العدم والبديلة) كما يلي:

فرضية العدم: لا يوجد فرق بين متوسطي المجتمعين: $H_0: \mu_1 - \mu_2 = 0$

الفرضية البديلة: وتشير إلى أنه يوجد فروق بينهما: $H_1: \mu_1 \neq \mu_2$

الحالة الأولى: وهي الحالة التي يكون فيها $\sigma_1^2 \neq \sigma_2^2$، وهي الحالة التي تشير إليها بيانات الجدول .

وفي هذه الحالة نقوم بحساب قيمة مؤشر الاختبار t من العلاقة العامة:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(2350.761 - 1535.488) - 0}{\sqrt{\frac{(1116.602)^2}{14} + \frac{(479.3964)^2}{42}}}$$

$$t = \frac{815.273}{\sqrt{89057.1447 + 5471.926388}} = \frac{815.273}{307.4558} = 2.65167$$

ولاتخاذ القرار حول H_0 نستخدم كلتا الطريقتين التاليتين:

- **طريقة القيمة الحرجة $t_{\frac{\propto}{2}}$:**

ولاتخاذ القرار المناسب بطريقة القيمة الحرجة $t_{\frac{\propto}{2}}$ حول الفرضية H_0 نبحث في الجداول الإحصائية لتوزيع (ستودينت) عن القيمة الحرجة $t_{\frac{\propto}{2}}$ المقابلة لدرجة حرية مساوية لأصغر العددين: $(n_1 - 1)$ أو $(n_2 - 1)$ فنجد أن $df = (n_1 - 1) = (14 - 1) = 13$، وبما أن الاختبار ثنائي الجانب, لأن $H_1: \mu_1 \neq \mu_2$ ، نجد أن القيمة الحرجة لـ t تساوي: $t_{\frac{\propto}{2},13} = t_{0.25\,,13} = 2.1604$

وبمقارنة t المحسوبة مع $t_{\frac{\propto}{2}}$ الحرجة نجد أن : $2.65167 > 2.1604$, أي أن $t > t_{\frac{\propto}{2}}$.

لذلك نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1 التي تقول أن $\mu_1 \neq \mu_2$. أي أن متوسط البروتين عند مرضى السرطان لا يساوي متوسطه عند الأشخاص الطبيعيين باحتمال ثقة(0.95) على الأقل.

ملاحظة: كان يمكننا الاستفادة من بيانات العينة ووضع الفرضية البديلة H_1 على الشكل الأحادي اليميني $H_1: \mu_1 > \mu_2$, وفي هذه الحالة يجب علينا أن نقارن t المحسوبة مع القيمة الحرجة $t_{13}(\propto)$ المقابلة لجانب واحد . لذلك نبحث في جداول (ستودينت) عن القيمة الحرجة $t_{13}(\propto)$ ، فنجد أن:

$$t_{13}(\propto) = t_{13}(0.05) = 1.7709$$

وبالمقارنة نجد أن: $t = 2.65167 > 1.7709$ لذلك نرفض فرضية العدم أيضاً . ونقبل بأن $\mu_1 > \mu_2$, أي نقبل بأن متوسط البروتين عند مرضى السرطان أكبر من متوسطه عند الأشخاص الطبيعيين باحتمال ثقة(0.95) على الأقل.

- **طريقة الاحتمال P:**

لاتخاذ القرار المناسب بطريقة احتمال الدلالة P، علينا أن نقوم بحساب قيمة P المقابلة للقيمة المحسوبة $t = 2.65167$, وبما أن الاختبار ثنائي فهي تساوي ضعف المساحة المحسوبة من تكامل توزيع (ستودينت) على المجال $]t\,, +\infty[$. وهذا يقتضي تحديد درجة الحرية الدقيقة المعرفة في توزيع (ستودينت) المستخدم في هذا الاختبار، وهناك عدة طرق لحساب الدرجة df وأهمها الطريقتان التاليتان:

- **الطريقة الدقيقة لحساب P**: وهي طريقة معقدة وتطبق في البرامج الحاسوبية، ولحساب درجة الحرية اللازمة تستخدم العلاقة الآتية [Triola, P.390]:

$$d\not{f} = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}} = 14.62907983 \qquad (30-6)$$

وبما أن قيمة $d\not{f}$ التي حصلنا عليها كانت على شكل عدد كسري (غير صحيح)، لذلك نقوم بحساب قيمتي P المقابلتين لدرجتي الحرية الصحيحتين المجاورتين للقيمة الفعلية $d\not{f}$ وهما 14 و15، فنجد من تكامل توزيع (ستودينت) على المجال $]2.65167, +\infty[$ أن:

$$P_{14} = 2 * (0.0094832) = 0.0189664 \qquad \text{(للجانبين)}$$

$$P_{15} = 2 * (0.0090654) = 0.0181308 \qquad \text{(للجانبين)}$$

ولحساب القيمة الحقيقية لـP المقابلة لدرجة الحرية الكسرية (14.629) نستخدم العلاقة التناسبية التالية:

$$P = P_{14} + (P_{15} - P_{14})(d\not{f} - 14)$$

$$P = 0.0189664 + (-0.0008356)(0.6290783)$$

$$P = 0.01844074$$

وهي قيمة قريبة جداً من قيمة P التي نحصل عليها من الحاسوب، وهذا يعني أن الحاسوب يتبع الطريقة الدقيقة والمعقدة في حساب P، وبما أن $P < \propto$ فإننا نرفض فرضية العدم H_0 التي تقول أن $\mu_1 = \mu_2$ ونقبل الفرضية البديلة التي تقوم أن $\mu_1 \neq \mu_2$، أي أن متوسط البروتين عند مرضى السرطان لا يساوي متوسطه الطبيعي باحتمال ثقة أكبر بكثير من (0.95) ، وهو يساوي:

$$1 - P = 1 - 0.01844 = 0.98156$$

ملاحظة: هناك من يختصر هذه الحسابات ويحسب درجة الحرية من العلاقة (df= $d\not{f}$ -2) ثم يقوم بحساب أو إيجاد قيمة P المقابلة لها كما هو مبين في الطريقة التقريبية التالية:

- **الطريقة التقريبية لحساب P:**

لحساب قيمة P التقريبية المقابلة لـ ($t = 2.65167$) في توزيع (ستودينت) نقوم بتحديد درجة الحرية $d\not{f}$ من أصغر العددين $(n_1 - 1)$ و$(n_2 - 1)$، فنجد أنها $(d\not{f} = 14 - 1 = 13)$، ;كما يمكن تقديرها من تعديل نتيجة الطريقة الدقيقة السابقة كما يلي: $d\not{f} = 14.62907983 - 2 = 13$.

ومن الجداول الإحصائية لقيم متحول (ستودينت)، نجد أن قيمة ($t = 2.65167$) المحسوبة والموافقة لدرجة حرية $d\not{f} = 13$ تجعل الاحتمال P (من الطرفين) يساوي:

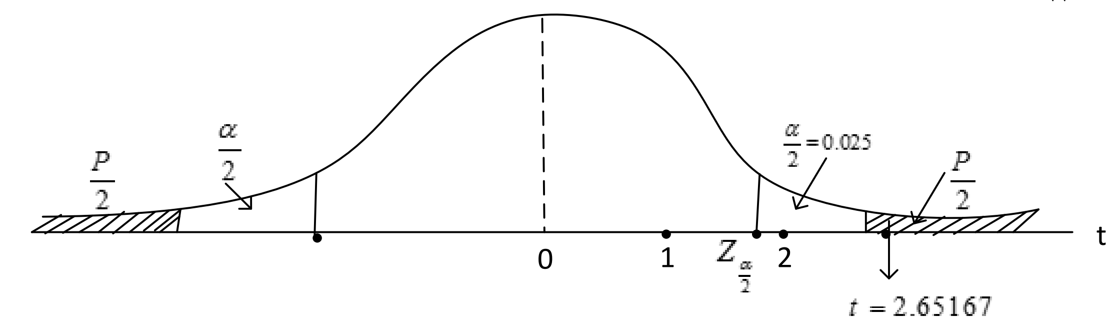
$$P = 0.0099741 * 2 = 0.0199482$$

وبما أن قيمة P أصغر من مستوى الدلالة $\propto = 0.05$, نرفض فرضية العدم H_0 التي تقول بعدم وجود فرق معنوي بين متوسطي المجتمعين، ونقبل الفرضية البديلة H_1 التي تقول أن متوسط المجتمع الأول

(مرضى السرطان) لا يساوي المتوسط الطبيعي. وإن ذلك موثوق باحتمال ثقة أكبر بكثير من (0.95) ، وهو يقترب من الواحد لأنه يساوي:

$$1 - P = 1 - 0.0199482 = 0.9800518 \approx 98\%$$

والشكل التالي يوضح معنى P بالمنطقة المظللة تحت منحنى التوزيع الطبيعي المعياري (المقارب لتوزيع (ستودينت)):



الشكل (6–12): تحديد المنطقة P

الحالة الثانية: وهي الحالة التي يكون فيها: $\sigma_1^2 = \sigma_2^2 = \sigma^2$، وعندها نقوم بحساب مؤشر الاختبار t من العلاقة (6– 24) التالية:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$t = \frac{(2350.761 - 1535.488) - 0}{\sqrt{\frac{13(1116.602)^2 + 41(479.3664)^2}{14 + 42 - 2}\left(\frac{1}{14} + \frac{1}{42}\right)}}$$

$$t = \frac{815.273}{212.60} = 3.8345$$

علماً بأن t يخضع تقاربياً لتوزيع (ستودينت) بدرجة حرية $n_1 + n_2 - 2$، ولهذا فإننا نقوم بحساب قيمة P المقابلة لـ $(t = 3.8345)$ ولدرجة حرية $(df = 14 + 42 - 2 = 54)$ ، فنجد أن الحاسوب يعطينا أن:

$$P = 2 * (0.00016548) = 0.00033096$$

وهي قيمة أصغر بكثير من مستوى الدلالة $\propto = 0.05$، لذلك نرفض فرضية العدم H_0 ونقبل الفرضية H_1 التي تقول بوجود فرق معنوي بين المتوسطين، ولكن هذه النتيجة محفوفة بالخطأ لأن الشرط المستخدم فيها $(\sigma_1^2 = \sigma_2^2 = \sigma^2)$ غير محقق في بيانات المثال المذكور، ولا يجوز الاعتماد عليها قبل إجراء اختبار لتساوي التباينين σ_1^2 و σ_2^2، ولقد قمنا بتطبيقها هنا للتدريب فقط.

ولاختبار الفرق بين متوسطي البروتين عند مرضى الربو والأشخاص الطبيعيين نتبع نفس الخطوات ونستخدم نفس العلاقات ونترك ذلك للقارئ على سبيل التدريب.

6-4-3: اختبار F لتساوي تبايني مجتمعين طبيعيين σ_2^2 و σ_1^2 :

لإجراء هذا الاختبار نضع فرضية العدم لتساوي التباينين H_0 على الشكل التالي:

$$H_0: \frac{\sigma_1^2}{\sigma^2} = 1 \qquad \Rightarrow H_0: (\sigma_1^2 = \sigma_2^2) \qquad (31-6)$$

ونضع الفرضية البديلة كما يلي:

$$H_1: \frac{\sigma_1^2}{\sigma_2^2} > 1 \qquad : \ (\sigma_1^2 > \sigma_2^2) \qquad (32-6)$$

ثم نسحب من المجتمعين عينتين عشوائيتين بحجمين n_2 و n_1، ونحسب تباينيهما المصححين s_2^2 و s_1^2.

ملاحظة: لتسهيل الحسابات تم تصميم جداول التوزيع F بحيث يكون رقم المجتمع الأول لصاحب التباين الأكبر (لذلك نرقم المجتمعين بحيث يكون $s_1^2 > s_2^2$ ، ونعدل الرموز في H_0 و H_1 حسب ذلك الترقيم) ثم نقوم بحساب مؤشر الاختبار F المعرف بالعلاقة التالية:

$$F = \frac{\dfrac{(n_1-1)S_1^2}{\sigma_1^2(n_1-1)}}{\dfrac{(n_2-1)S_2^2}{\sigma_2^2(n_2-1)}} = \frac{\dfrac{S_1^2}{\sigma_1}}{\dfrac{S_2^2}{\sigma_2^2}} = \frac{\sigma_2^2}{\sigma_1^2} * \frac{S_1^2}{S_2^2} \sim F_{v_1 v_2} \qquad (33-6)$$

ولكن بما أن فرضية العدم تنص على أن: $\sigma_1^2 = \sigma_2^2$ فإن مؤشر الإختبار F يختصر ويأخذ الشكل التالي:

$$F = \frac{S_1^2}{S_2^2} \qquad : (S_1^2 > S_2^2) \qquad (34-6)$$

وهو متحول عشوائي يخضع لتوزيع فيشير $F_{(x)}$ بدرجتي حرية $v_1 = (n_1 - 1)$ للبسط و$v_2 = (n_2 - 1)$ للمقام، وبعد حساب قيمة F نقارنها مع القيمة الحرجة $F_{(\propto)}$ (لاتجاه واحد) والمقابلة لكامل $\propto$ ولدرجتي الحرجة $v_1 = (n_1 - 1)$ و$v_2 = (n_2 - 1)$ ونتخذ القرار كما يلي :

إذا كانت $F \leq F_{(\propto)}$ (أو كانت $P > \propto$) فإننا نقبل الفرضية H_0 التي تقول بتساوي التباينين $\sigma_1^2 = \sigma_2^2$.

أما إذا كانت $F > F_{(\propto)}$ نرفض H_0 ونقبل H_1 ونقول بعدم تساوي التباينين المذكورين .

6-5: اختبارات معالم عدة مجتمعات طبيعية (من عدة عينات مستقلة):

6-5-1: اختبار تساوي متوسطات عدة مجتمعات طبيعية :

يطبق هذا الاختبار لمقارنة المتوسطات في أكثر من مجتمعين طبيعيين (3 فأكثر)، ولذلك نفترض أن عدد المجتمعات هو g $\left(g > 2\text{حيث}\right)$، وإننا سحبنا منهم عشوائياً عينات مستقلة بحجوم n_1 و n_2 و n_3 ... n_g ، وكانت متوسطاتها $\bar{x}_1$ و $\bar{x}_2$ و $\bar{x}_3$ ... $\bar{x}_g$ وكان متوسط المتوسطات هو $\bar{\bar{x}}$.

وكانت تبايناتها المصححة s_1^2 و s_2^2 و s_3^2 ... s_g^2

فإننا نضع فرضيتي العدم والبديلة حول متوسطات هذه المجتمعات كما يلي:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_g$$

$$H_1: \mu_k \neq \mu_J \quad \text{لزوج } (k, \text{J}) \text{ واحد على الأقل :} \qquad (35-6)$$

كما نفترض أن تباينات هذه المجتمعات متساوية وتساوي σ^2 .

ثم نحسب مجاميع مربعات الانحرافات المختلفة وهي:

مجموع مربعات الانحرافات داخل العينات، أي مربعات (الخطأ) :

$$SSE = \sum_{k=1}^{g} (n_k - 1) s_k^2 \qquad (36-6)$$

حيث g: عدد المجتمعات.

مجموع مربعات الانحرافات بين العينات:

$$SSB = \sum_{k=1}^{g} n_i (\bar{x}_k - \bar{\bar{x}})^2 \qquad (37-6)$$

مجموع مربعات الانحرافات الكلية لجميع عناصر العينات:

$$SST = \sum_{k=1}^{g} \sum_{i=1}^{n_k} (x_{ki} - \bar{\bar{x}})^2 = \sum_{k=1}^{g} \sum_{i=1}^{n_k} x_{ki}^2 - \frac{T^2}{n} \qquad (38-6)$$

حيث أن: $n = \sum n_k$ وأن: $T = \sum_{k=1}^{g} \sum_{i=1} x_{ki}$

ويمكن البرهان عل أنه يكون لدينا:

$$SST = SSB + SSE \qquad (39-6)$$

علماً بأن درجات الحرية لكل منهم، هي و$(n-g)$ و$(g-1)$ و$(n-1)$ على الترتيب، ثم نضع النتائج في جدول كالتالي:

جدول(6-3) : نتائج تحليل التباين الأحادي ANOVA

مصدر التباين	مجموع المربعات ورمزه	درجة الحرية	متوسط المربعات	قيمة F المحسوبة	قيمة F الحرجة	قيمة P
التباين بين العينات	SSB	$g-1$	$MSSB = \frac{SSB}{g-1}$	$F = \frac{MSSB}{MSSE}$	$F_\propto$	P
التباين داخل العينات	SSE	$n-g$	$MSSE = \frac{SSE}{n-g}$	ــــــ	ــــــ	ــــــ
التباين الكلي	SST	$n-1$	ــــــ	ــــــ	ــــــ	ــــــ

ثم نحسب قيمة مؤشر الاختبار F المعرف بالعلاقة التالية :

$$F = \frac{MSSB}{MSSE} \qquad (40-6)$$

وهو يخضع للتوزيع F بدرجتي حرية $(g-1\ ,n-g)$، ونتعامل معه كما تعاملنا مع F السابقة عند اتخاذ القرار حول H_0 في اختبار تساوي التباينين . فإذا كان $F \le F(\propto)$ نقبل H_0 والعكس بالعكس .

ملاحظة: يُسمى هذا الاختبار بتحليل التباين ANOVA باتجاه واحد (Analysis Of Variance- one way) وسنقوم بدراسته بالتفصيل في الفصل السابع.

مثال (6–4): [مأخوذ من Copal P.56 **بتصرف]**

لنفترض أنه لدينا (3) مجتمعات طبيعية، ونريد اختبار تساوي متوسطات متحول واحدX فيها, وبمستوى دلالة $\propto = 0.05$، لذلك سحبنا (3) عينات عشوائية منها بحجوم: $n_1 = 3 \quad n_2 = 5 \quad n_3 = 4$,

ووضعنا فرضيتي العدم والبديلة حول متوسطات هذه المجتمعات كما يلي:

$$H_0: \mu_1 = \mu_2 = \mu_3$$

من أجل(k,J)واحد على الأقل : $\quad H_1: \mu_k \neq \mu_J$

ولنفترض أن البيانات الأصلية (غير الموجودة) لهذه العينات، أعطتنا أن مجاميع قياسات X فيها كانت تساوي ما يلي:

$$\sum_{i=1}^{3} x_{1i} = 53.5 \quad , \quad \sum_{i=1}^{5} x_{2i} = 102.5 \quad , \quad \sum_{i=1}^{4} x_{3i} = 64.4$$

وإن مجموعها الكلي يساوي:

$$T = \sum_{k=1}^{3}\sum_{i=1}^{n_k} x_{ki} = 53.5 + 102.5 + 64.4 = 220.4$$

وإن متوسطات X في العينات المسحوبة تساوي:

$$\bar{x}_1 = \frac{53.5}{3} = 17.83 \quad , \quad \bar{x}_2 = \frac{102.5}{5} = 20.50 \quad , \quad \bar{x}_3 = \frac{64.4}{4} = 16.10$$

وإن المتوسط العام لـ X فيها (أو المتوسط المثقل للمتوسطات) يساوي:

$$\bar{x} = \frac{T}{\sum n_i} = \frac{220.4}{12} = 18.37$$

ثم نقوم بحساب الكسر $\frac{T^2}{n}$ فنجد أن:

$$\frac{T^2}{n} = \frac{(220.4)^2}{12} = 4048.01$$

ثم نقوم بحساب SST من العلاقة (6–38) فنجد من البيانات الأصلية أن:

$$SST = \sum_{k=1}^{3}\sum_{i=1}^{n_k} x_{ki}^2 - \frac{T^2}{n} = [(954.43 + 2105.13 + 1037.98) - 4048.01]$$

$$SST = 4097.54 - 4048.1 = 49.53$$

ثم نقوم بحساب SSB من العلاقة (6−37) فنجد أن:

$$SSB = \sum_{k=1}^{3} x_k(\bar{x}_k - \bar{x})^2 = 3(17.83 - 18.37)^2 + 5(20.50 - 18.37)^2 + 416.10 - 18.37^2$$
$$SSB = 44.17$$

ثم نقوم بحساب SSE من العلاقة:

$$SSE = SST - SSB = 49.53 - 44.17 = 5.36$$

ثم نقوم بوضع نتائج هذه الحسابات في جدول كالتالي:

جدول (6−4): جدول ANOVA

مصدر التباين	مجموع المربعات	درجة الحرية	متوسطات المربعات	
بين العينات SSB	$SSB = 44.17$	$g - 1 = 2$	$MSSB = 22.09$	$F = \frac{22.09}{0.556} = 37$
داخل العينات SSE	$SSE = 5.36$	$n - g = 9$	$MSSE = 0.556$	———
التباين الاجمالي SST	$SST = 49.53$	$n - 1 = 11$	———	———

ثم نقوم بحساب قيمة المؤشر F من العلاقة:

$$F = \frac{\frac{SSB}{g-1}}{\frac{SSE}{n-g}} = \frac{\frac{44.17}{2}}{\frac{5.36}{9}} = \frac{22.09}{0.556} = 37$$

ولمقارنة قيمة F المحسوبة مع قيمتها الحرجة $F_{v_1,v_2}(\propto)$ المقابلة لمستوى الدلالة $(\propto = 0.05)$ ولدرجتي الحرية $v_1 = g - 1 = 2$, $v_2 = n - g = 9$ ، علينا أن نبحث عن قيمة $F_{v_1,v_2}(\propto)$ في جداول F فنجد أنها تساوي :

$$F_{v_1,v_2}(\propto) = F_{2,9}(0.05) = 4.24$$

ثم نقوم بمقارنة F المحسوبة مع $F_{2,9}(0.05)$ الحرجة، فنجد أن $F > F_{2,9}(0.05)$، لذلك نرفض فرضية العدم H_0، التي تقول أن $H_0: \mu_1 = \mu_2 = \mu_3$، ونقبل الفرضية H_1 التي تقول أن أحد متوسطات هذه المجتمعات (على الأقل) يختلف عن الأخرى. (ولعله المجتمع الثاني لأن $\bar{x}_2 = 20.50$).

6−5−2: اختبار تساوي تباينات عدة مجتمعات طبيعية (من عدة عينات مستقلة):

6−5−2−1: اختبار (بارتليت Bartlett test) لتساوي تباينات عدة مجتمعات:

لنفترض إننا نريد دراسة تباينات متحول X في g مجتمعاً طبيعياً أو شبه طبيعي $(g > 2)$. لذلك سحبنا منها g عينة عشوائية بحجوم مختلفة أو متساوية نرمز لها بـ $n_1, n_2, \dots n_k \dots n_g$، ولمجموع حجومها بـ $n = \sum_{k=1}^{n} n_k$. ثم نقوم بحساب متوسطاتها ورمزنا لها بـ $\bar{x}_1, \bar{x}_2, \dots \bar{x}_k \dots \bar{x}_g$، وحساب تبايناتها ورمزنا لها بـ $s_1^2, s_2^2, \dots s_k^2, \dots s_g^2$، والآن لنفترض أن تباينات X في هذه المجتمعات

هي :

$$\sigma_1^2\ ,\sigma_2^2\ , \dots \sigma_k^2\ , \dots \sigma_g^2$$

ثم نضع الفرضيتين حول تساوي هذه التباينات كما يلي:

$$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \cdots = \sigma_g^2 \qquad (41-6)$$

$H_1: \sigma_k^2 \neq \sigma_\ell^2$: من أجل زوج واحد $(k\ ,\ell)$ على الأقل

أي أننا نفترض في H_0 أن تباينات X في هذه المجتمعات متساوية، مقابل الفرضية البديلة H_1 التي تعني أنها غير متساوية من أجل مجتمعين على الأقل.

ولاختبار هذه الفرضية قام Bartlett باستخراج مؤشر خاص وعرفه بالعلاقة التالية:

$$BT = \frac{(n-g)\ ln\, S_p^2 - \sum_{k=1}^{g}(n_k-1)\ ln\, s_k^2}{1+\left[\frac{1}{3(g-1)}\right]\left[\sum_{k=1}^{g}\frac{1}{n_k-1}-\frac{1}{n-g}\right]} \sim \chi_{g-1}^2 \qquad (42-6)$$

حيث أن: n هو حجم العينة الكلية $n = \sum_{k=1}^{n} n_k$، و g عدد المجتمعات.

وحيث أن: s_k^2 هو تباين X في العينة ، وأن: S_p^2 هو التباين المدمج المحسوب من العلاقة :

$$S_p^2 = \frac{\sum(n_k-1)s_k^2}{n-g} \qquad (43-6)$$

وبرهن على أن هذا المؤشر يخضع تقاربياً لتوزيع χ_{g-1}^2 بدرجة حرية $(g-1)$.

لذلك فإننا عند اتخاذ القرار حول الفرضية H_0 نقارن القيمة المحسوبة BT مع القيمة الحرجة $\chi_{g-1}^2(\propto)$ ونتخذ القرار عند مستوى دلالة ∝ كما يلي:

(44–6) إذا كانت $BT \le \chi_{g-1}^2(\propto)$، نقبل الفرضية H_0 والتي تنص على أن التباينات متساوية,

أما إذا كان $BT > \chi_{g-1}^2(\propto)$ فإننا نرفض H_0 ونقبل الفرضية البديلة H_1، التي تنص على أن التباينات غير متساوية في مجتمعين على الأقل.

مثال (6–5): لنفترض أنه لدينا (5) خطوط لعصر الزيتون (معاصر) ونريد دراسة فيما إذا كانت تباينات الانتاج اليومي فيها متساوية أم مختلفة . لذلك وضعنا الفرضيتين كما يلي:

$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2$$

$H_1: \sigma_k^2 \neq \sigma_\ell^2$: من أجل خطين على الأقل

ثم قمنا بسحب (5) عينات طبقية من إنتاج هذه الخطوط خلال أربعة أيام $(n_k = 4)$ فحصلنا منها على البيانات التالية:

جدول (6–5): كميات الإنتاج اليومية (بالكغ) حسب الخطوط والأيام

المجموع	الخط E	الخط D	الخط C	الخط B	الخط A	الخطوط / رقم اليوم	
	250	340	250	310	250	1	
	240	270	230	330	260	2	
	270	300	220	280	230	3	
	290	320	260	360	270	4	
$n = 20$	4	4	4	4	4	**الأعداد**	n_k
1382.5	262.5	307.5	240.0	320.0	252.5	**المتوسطات**	$\bar{x}_k$
3141.662	491.667	891.667	333.33	1133.33	291.667	**التباينات**	s_k^2
31.5094	6.1978	6.7931	5.8091	7.0329	5.6756	**لوغاريتمات التباينات**	$ln\, s_k^2$

ولمتابعة الحل قمنا بحساب بعض القيم الاحصائية لتلك البيانات ووضعنا في أسفل الجدول السابق.

والآن نقوم بحساب الكميات التي تدخل في تعريف الاختبار BT فنجد أن التباين المدمج S_p^2 يساوي (انظر الجدول السابق):

$$S_p^2 = \frac{\sum^g (n_k - 1)s_k^2}{n - g} = \frac{3(\sum s_k^2)}{20 - 5} = \frac{3 * (3141.667)}{15}$$

$$S_p^2 = \frac{9425}{15} = 628.333$$

كما نجد أن الحد الثاني في البسط يساوي:

$$\sum_{k=1}^{g} (n_k - 1)\, ln\, s_k^2 = 3\left(\sum_{k=1}^{g} ln\, s_k^2\right) = 3(31.5094) = 94.5292$$

ثم نقوم بحساب المقام فنجد أنه يساوي:

$$C = 1 + \left(\frac{1}{3(5-1)}\right)\left[\left(\frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3}\right) - \left(\frac{1}{20 - 5}\right)\right]$$

$$C = 1 + \frac{1}{12}\left[\frac{5}{3} - \frac{1}{15}\right] = 1.1333$$

نعوض نتائج هذه الحسابات في معادلة المؤشر BT فنجد أن:

$$BT = \frac{(20-5)\, ln(628.333) - 945292}{1.1333} = \frac{2.1167}{1.1333} = 1.8678$$

ثم نقوم بإيجاد القيمة الحرجة لـ $\chi_{g-1}^2(\propto)$ عند مستوى الدلالة $\propto = 0.05$ ودرجة الحرية المساوية لـ $g - 1 = 5 - 1 = 4$, فنجد أن:

$$\chi_{g-1}^2(\propto) = \chi_4^2(0.05) = 9{,}488$$

وبمقارنة القيمة المحسوبة للمؤشر BT مع القيمة الحرجة $\chi_4^2(\propto)$, نجد أن $1.8678 < 9.488$, لذلك نقبل فرضية العدم H_0 التي تقول أن تباينات الانتاج على تلك الخطوط متساوية وباحتمال ثقة 0.95 .

ملاحظة: بعد إن تأكدنا من تساوي التباينات , يمكننا أيضاً دراسة تساوي متوسطات الانتاج على هذه الخطوط وعندها يجب أن نستخدم الاختبار F وإجراء تحليل التباين ANOVA كما فعلنا في المثال (6−4) السابق، ونترك ذلك للقارئ على سبيل التمرين.

6−5−2−2: اختبار (ليفيني Levene) لتساوي التباينات في عدة مجتمعات (من عدة عينات مستقلة)

يستخدم هذا الاختبار لدراسة تساوي أو تجانس تباينات متحول X في عدة مجتمعات طبيعية، وهو يقدم لنا خدمة جليلة عند تطبيق الكثير من الاختبارات الإحصائية، التي تفترض أن تباينات X في المجتمعات المدروسة متساوية، لأنه يساعدنا على التحقق من صحة تلك الافتراضات، ويعتبر هذا الاختبار بديلاً لاختبار Bartlett، ولكنه أقل حساسية منه في الاعتماد على التوزيع الطبيعي.

فإذا كان لدينا شك قوي بأن البيانات المستخدمة ليست مسحوبة من مجتمع طبيعي (أو شبه طبيعي) فإنه يفضل استخدام اختبار Bartlett ، لأنه يعطينا نتائج أفضل منه.

ولإجراء هذا الاختبار نفترض أننا نريد اختبار تساوي تباينات متحول طبيعي X في عدة مجتمعات (أو مجموعات)، ولنفترض أن عدد تلك المجتمعات g $\left(\text{حيث } g > 2\right)$ وسحبنا منها g عينة عشوائية بحجوم مختلفة أو متساوية: $n_1 , n_2 , \dots\ n_k \dots\ n_g$ حيث $(n = \sum n_k)$، وحصلنا منها على متوسطاتها: $\bar{x}_1 , \bar{x}_2 , \dots\ \bar{x}_k \dots\ \bar{x}_g$ ، وعلى تبايناتها التالية: $s_1^2 , s_2^2 , \dots s_k^2 , \dots s_g^2$. والآن لنفترض أن تباينات X في تلك المجتمعات هي: $\sigma_1^2 , \sigma_2^2, \dots , \sigma_k^2, \dots \sigma_g^2$.

وبناء على ذلك نضع الفرضيتين الاحصائيتين كما يلي:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \cdots = \sigma_g^2 \qquad (45-6)$$

$H_1 : \sigma_k^2 \neq \sigma_\ell^2$: من أجل زوج واحد (k , ℓ) على الأقل

أما مؤشر الاختبار فيعرف حسب Levene بالعلاقة التالية:

$$W = \frac{(n-g)}{(g-1)} \frac{\sum_{k=1}^{g}(\bar{Z}_k - \bar{Z})^2}{\sum\sum(Z_{ki} - \bar{Z}_k)^2} \quad : \left(n = \sum n_k\right) \qquad (46-6)$$

حيث أن المتحول Z هو تحويل من المتحول X وفق إحدى العلاقات الثلاثة التالية:

$1-\quad Z_{ki} = |x_{ki} - \bar{X}_k|$: حيث أن : $\bar{X}_k$ متوسط X في العينة K $\qquad (47-6)$

$2-\quad Z_{ki} = |x_{ki} - X'_k|$: أو حيث أن : X'_k وسيط X في العينة K $\qquad (48-6)$

$3-: Z_{ki} = |x_{ki} - X''_k|$: أو أن : X''_k هو المتوسط المرتب لـ 10% الأولى من قيم X $\qquad (49-6)$

أما متوسطات Z فتحسب كما يلي:

$$\bar{Z}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} Z_{ki}$$: K متوسط القيم Z_{ki} في العينة (50 − 6)

$$\bar{Z} = \frac{1}{g}\sum_{i=1}^{n_k} \bar{Z}_k = \frac{1}{n}\sum_{k=1}^{g}\sum_{i=1}^{n_k} Z_{ki}$$: المتوسط العام لـ Z_{ki} (51 − 6)

وهنا نشير إلى أن التعاريف الثلاثة لـ Z_{ki} تساعدنا في تحديد حصانة وقوة اختبار (ليفيني)، ويقصد بمصطلح الحصانة قدرة الاختبار على عدم إعطاء إشارة مزيفة عن عدم تساوي التباينات, عندما تكون البيانات غير خاضعة للتوزيع الطبيعي وتكون التباينات فعلياً متساوية، ويقصد بالقوة قدره الاختبار على اكتشاف التباينات غير المتساوية عندما تكون التباينات فعلياً غير متساوية.

وأخيراً نشير إلى أن مؤشر الاختبار W يخضع لتوزيع F بدرجتي حرية $v_1 = g - 1$ و $v_2 = \mathrm{n} - g$ (حيث أن: $n = \sum_{k=1}^{g} n_k$).

ولاتخاذ القرار حول نتيجة الاختبار نقارن قيمة W المحسوبة من العلاقة (6−46) بالقيمة الحرجة $F_{v_1, v_2}(\propto)$ ونتخذ القرار كما يلي:

إذا كانت $W \le F_{v_1, v_2}(\propto)$ نقبل فرضية العدم H_0

أما إذا كانت $W > F_{v_1, v_2}(\propto)$ نرفض H_0 ونقبل الفرضية H_1 $(51a - 6)$

مثال (6−6): لنفترض أنه لدينا (10) مجموعات من الطلاب، ونريد اختبار تساوي تباينات أعمارهم في تلك المجموعات، فسحبنا من كل مجموعة عينة عشوائية بحجوم متساوية: $(n_k = 5)$ طلاب، فكان حجم العينة الكلية $n = 50$ طالباً . وبعد أخذ بيانات الأعمار في كل مجموعة وحساب متوسطاتها $\bar{x}_k$ وضعنا فرضيتي الاختبار كما يلي:

$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 \ldots = \sigma_{10}^2$$

$$H_1: \sigma_k^2 \neq \sigma_\ell^2$$: من أجل زوج واحد على الأقل

ثم قمنا بإجراء التحويلات من X إلى Z حسب إحدى العلاقات السابقة، ولتكن العلاقة (6−46) المستندة إلى المتوسطات $\bar{x}_k$، فنحصل على القيم Z_{ki}، ثم نقوم بحساب المتوسطات $\bar{Z}_k$ وأخيراً نقوم بحساب قيمة مؤشر الاختبار W ولنفترض أنها كانت تساوي:

$$W = \frac{(50-10)}{(10-1)} \frac{\sum 5(\bar{Z}_k - \bar{Z})^2}{\sum\sum(Z_{ki} - \bar{Z}_k)^2} = 1.75$$

وبما أن درجتي الحرية تساويان $v_1 = 10 - 1 = 9$ و$v_2 = 50 - 10 = 40$, فإننا نقارن القيمة المحسوبة W مع القيمة الحرجة $F_{v_1, v_2}(\propto)$. وباعتبار أن $\propto = 0.05$ نجد من جداول F أن:

$$F_{v_1, v_2}(\propto) = F_{9, 40}(0.05) = 2.124$$

وبمقارنة القيمة المحسوبة W مع $F_{9, 40}(0.05)$ نجد أن: $1.75 \le 2.124$, لذلك نقبل فرضية العدم

H_0 التي تقول أن تباينات العمر X في هذه المجموعات متساوية وذلك باحتمال ثقة 0.95 على الأقل .

وبما أن التباينات متساوية فإننا نقوم بمقارنة الفروقات بين المتوسطات باستخدام إحدى العلاقتين (6−40) أو (6−25) .

6−6: اختبار الأزواج المتقابلة (من عينتين مرتبطتين):

يطبق هذا الاختبار لمقارنة نتائج إجابات أو علامات عينية مؤلفة من نفس الأشخاص, قبل التجربة وبعدها، لذلك تسمى الدرجات الأولى بالدرجات القبلية وتسمى الدرجات الثانية بالدرجات البعدية، ويمكن وضع النتائج في جدول خاص على شكل أزواج متقابلة (كل زوج لشخص واحد) فنحصل على عينتين مرتبطتين من تلك الدرجات كما يلي:

جدول (6−6): البيانات المتقابلة

رقم الشخص	1	2	3	4	.....	i	.....	n	المتوسط	
الدرجات القبلية	x_1	x_2	x_3	x_4	.....	x_i	.....	x_n	$\bar{x}$	
الدرجات البعدية	y_1	y_2	y_3	y_4	.....	y_i	.....	y_n	$\bar{y}$	
الفروقات d_i	d_1	d_2	d_3	d_4	.....	d_i	.....	d_n	$\bar{d}$	S_d

ثم نقوم بحساب الفروقات بين قيمتي كل زوج من العلاقة : $d_i = x_i - y_\text{i}$

ثم نقوم بحساب متوسط وتباين هذه الفروقات من العلاقتين :

$$\bar{d} = \frac{1}{n}\sum d_i \qquad (52-6)$$

$$S_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2 = \frac{1}{n-1}\left[\sum d_i^2 - n\bar{d}^2\right] \qquad (53-6)$$

ثم نحسب الانحراف المعياري للفروقات d_i ونرمز له بـ S_d من العلاقة: $S_d = \sqrt{S_d^2}$.

ولإجراء هذا الاختبار حول الفرق بين العينتين نضع الفرضيتين كما يلي:

$$H_0: \bar{D} = 0 \quad , \quad H_1: \bar{D} > 0$$ (الاختبار أحادي يميني)

حيث $\bar{D}$: هو متوسط الفروقات في المجتمع ، وتؤخذ قيمته الصفرية من فرضية العدم H_0 .

ثم نحسب مؤشر الاختبار المعرف بالعلاقة:

$$t = \frac{\bar{d} - \bar{D}}{S_d/\sqrt{n}} \qquad (54-6)$$

ثم نقارن قيمة t المحسوبة مع قيمة t الحرجة $t_{\propto,n-1}$، فإذا كانت $t < t_{\propto,n-1}$ تقبل فرضية العدم

H_0، ونقول بأنه لا يوجد فرق بين الدرجات القبلية والبعدية، والعكس بالعكس.

ملاحظة1 : إذا كانت الفرضية البديلة من الشكل $H_1 : \overline{D} < 0$ فإننا سنحصل على قيمة سالبة لـ t ، لذلك نقارنها مع $(_\ t_{\propto,n-1})$، فإذا كانت $t < -t_{\propto,n-1}$ فإننا نرفض فرضية العدم H_0 ، والعكس بالعكس.

ملاحظة 2: إذا كان حجم العينة n>30 , فإننا نستبدل مؤشر الاختيار السابق t باختبار(ويلكوكسن Wilcoxon) المعرف *بالعلاقة:*

$$Z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \qquad (55-6)$$

وحيث أن T هو أصغر المجموعين التالين :مجموع القيم المطلقة للرتب السالبة للفروقات T^- , أو مجموع مجموع قيم الرتب الموجبة للفروقات T^+ (انظر الاختبارات اللامعلمية في الفصل العاشر).

مثال (6−7): لدراسة تأثير أحد الأدوية على مستوى ضغط الدم عند المرضى المصابين به. قرر أحد الباحثين إجراء تجربة هذا الدواء على (8) مرضى. ولذلك قام أولاً بقياس مستويات الضغط عند هؤلاء المرضى قبل إعطائهم الدواء، ثم قام بإعطائهم الدواء وبعد مرور ساعة على ذلك أخذ قياسات مستويات الضغط لهم، فحصل على البيانات القبلية والبعدية، ثم قام بحساب الفروقات الزوجية d_i وقام بتربيعها فحصل على الجدول التالي:

جدول (6−8)**: بيانات المثال.**

رقم المريض	1	2	3	4	5	6	7	8	المجموع
مستوى الضغط قبل التجربة X	170	175	180	175	160	170	175	180	——
مستوى الضغط بعد التجربة Y	150	160	170	160	170	160	170	185	——
الفروقات الزوجية d_i	20	15	10	15	−10	10	5	−5	+60
d_i^2	400	225	100	225	100	100	25	25	1200

ثم قام بحساب متوسط تلك الفروقات فوجد أن:

$$\bar{d} = \frac{\sum_{i=1}^{8} d_i}{n} = \frac{+60}{8} = 7.5$$

ثم قام بحساب تباين تلك الفروقات فحصل على أن :

$$S_d^2 = \frac{1}{n-1}\left[\sum d_i^2 - n\bar{d}^2\right] = \frac{1}{7}[1200 - 8(7.5)^2] = 107.14$$

$$S_d = \sqrt{S_d^2} = \sqrt{107.14} = 10.35$$

ثم قام بوضع فرضيتي الاختبار كما يلي:

$H_0: \bar{D} = 0 \quad , \quad H_1: \bar{D} > 0$ (الاختبار أحادي يميني)

ثم قام بحساب قيمة مؤشر الاختبار من العلاقة :

$$t = \frac{\bar{d} - \bar{D}}{S/\sqrt{n}} = \frac{7.5 - 0}{10.35/\sqrt{8}} = 2.05$$

وباعتماد $\propto = 0.05$ قام بمقارنة هذه القيمة المحسوبة مع القيمة الحرجة $t_{n-1}(\propto)$ والتي تساوي: $t_{n-1}(\propto) = t_7(0.05) = 1.895$، فوجد أن: $t > t_7(0.05)$، لذلك رفض فرضية العدم H_0 التي تقول أن $(\bar{D} = 0)$ وتم قبول الفرضية البديلة H_1 التي تقول أن $(\bar{D} > 0)$. وهذا يعني أن متوسط الفروقات كان موجباً، وهو ما يؤكد أن الدواء المستخدم له تأثير إيجابي على مستويات الضغط عند المرضى وباحتمال ثقة قدره 0.95 .

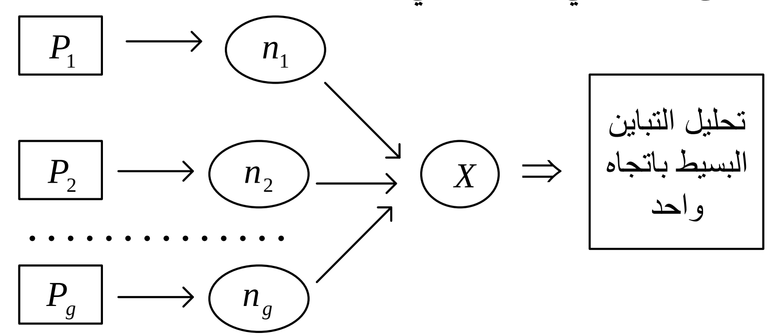
الفصل السابع
تحليل التباين البسيط (ANOVA)

سنتناول في هذا الفصل عدة أنواع من تحليل التباين البسيط هي:

- تحليل التباين باتجاه واحد (one-way ANOVA)
- تحليل التباين باتجاهين (two-way ANOVA)
- تحليل التباين بثلاث اتجاهات (three-way ANOVA)
- تحليل المربع اللاتيني (LATIN SQUARE)
- تحليل التباين المشترك باتجاه واحد (ANCOVA one way)

7-1: تحليل التباين البسيط باتجاه واحد (one-way ANOVA):

يتناول تحليل التباين البسيط باتجاه واحد دراسة تغيرات متحول واحد X (يسمى بالتابع), الناتجة عن عدة مجتمعات (أو معالجات) نرمز لها بـ $P_1\ P_2\ ...\ P_g$. وهنا يشترط أن يكون عدد المجتمعات $g > 2$ (لأنه إذا كان $g = 2$ فإننا نستخدم اختبار (ستودينت) t لمقارنة متوسطي المجتمعين), ويمكن تمثيل تأثير هذه المجتمعات على X كما في الشكل التالي:



الشكل (7-1): تمثيل ANOVA باتجاه واحد

ولدراسة تغيرات X الناتجة عن تأثيرات هذه المجتمعات نسحب من كل مجتمع k، عينة عشوائية بحجم n_k ، ثم نأخذ قياسات X من عناصر هذه العينات، ونضعها في جدول مناسب، يتضمن قياسات X من كل عينة n_k ومتوسطها $\bar{X}_k$ وتوقعها الرياضي في المجتمع μ_k , كما يمكن أن يتضمن تباينها S_k^2 كالجدول التالي:

جدول (7−1): قيم X حسب عينات المجتمعات

المجتمعات	حجوم العينات	قياسات X من عناصر العينات	متوسطات العينات	التوقعات في المجتمعات	تباينات العينات
P_1	n_1	x_{11} x_{12} x_{13} … x_{1n_1}	$\bar{X}_1$	μ_1	S_1^2
P_2	n_2	x_{21} x_{22} x_{23} … x_{2n_2}	$\bar{X}_2$	μ_2	S_2^2
⋮	⋮	⋮	⋮	⋮	⋮
P_K	n_k	x_{k1} x_{k2} x_{k3} … x_{kn_k}	$\bar{X}_K$	μ_k	S_K^2
⋮	⋮	⋮	⋮	⋮	⋮
P_g	n_g	x_{g1} x_{g2} x_{g3} … x_{gn_g}	$\bar{X}_g$	μ_g	S_k^2

ويشترط في هذه العينات والبيانات أن تحقق الشروط أو الافتراضات التالية:

1− أن تكون العينات المسحوبة من المجتمعات عشوائية ومستقلة عن بعضها البعض.

2− أن يكون تباين X في جميع هذه المجتمعات موحداً ويساوي σ^2 .

3− أن تكون قيم X في كل مجتمع k خاضعة للتوزيع الطبيعي $N(\mu_K ,\sigma^2)$ ، الذي توقعه μ_k وتباينه ثابت ويساوي σ^2 في كل المجتمعات.

ويمكننا تجاهل الشرط الثالث (حول الطبيعية) عندما تكون حجوم العينات كبيرة، وذلك بالاستناد على مفعول نظرية النهاية المركزية في الاحتمالات، والتي تنص على أن توزيعات X في تلك المجتمعات تنتهي إلى التوزيع الطبيعي.

وعندها فإن فرضية العدم والفرضية البديلة حول توقعات هذه المجتمعات تأخذان الشكل التالي:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k = \mu_g \qquad (1-7)$$
$$H_1 : \mu_k \neq \mu_\ell$$

وذلك من أجل زوج واحد على الأقل $(K \neq \ell)$

وإذا رمزنا للمتوسط المثقل لهذه التوقعات بالرمز $\bar{\mu}$ والذي يحسب من العلاقة التالية :

$$\bar{\mu} = \frac{n_1\mu_1 + n_2\mu_2 + \cdots + n_g\mu_g}{n_1 + n_2 + \cdots + n_g} = \frac{\sum^g n_k\mu_k}{\sum_{k=1} n_k} = \frac{\sum^g n_k\mu_k}{n} \qquad (2-7)$$

حيث أن: n_K هو حجم العينة المسحوبة من المجتمع k، وأن n مجموعها ، أي أن: $n = \sum_{K=1}^{g} n_k$,

ويُطلق على المتوسط $\bar{\mu}$ مصطلح المتوسط أو التوقع الكلي (Grand mean)، وبناء على ذلك يمكننا التعبير عن قيمة أي توقع μ_k بدلالة التوقع الكلي $\bar{\mu}$ كما يلي :

$$\mu_k = \bar{\mu} + (\mu_k - \bar{\mu}) \qquad (3-7)$$

أو على الشكل التالي :

$$\mu_k = \bar{\mu} + \tau_k \qquad (4-7)$$

حيث أن:

$$\tau_K = \mu_k - \bar{\mu} \qquad (5-7)$$

ونعبر عن ذلك لفظياً كما يلي:

(تأثير المجتمع k (المعالجة k)) + (التوقع الكلي) = (توقع X في المجتمع k)

وهذا يقودنا إلى تعديل فرضية العدم حول التوقعات μ_k إلى فرضية عدم جديدة H_0 مقابل H_1 كما يلي:

$$H_0 : \tau_1 = \tau_2 = \tau_3 = \cdots = \tau_g = 0 \qquad (6-7)$$

$$H_1 : \tau_k \neq 0$$ وذلك من أجل مجتمع واحد على الأقل.

وهذا يجعلنا نعتبر أن قياسات X في المجتمع k ، والتي سنرمز لها بـ x_{ki} ، خاضعة للتوزيع الطبيعي $N[(\bar{\mu} + \tau_k) , \sigma^2]$ وبالتالي يمكننا كتابة كل قياس منها x_{ki} كما يلي:

$$x_{ki} = \bar{\mu} + (\mu_k - \bar{\mu}) + (x_{ki} - \mu_k) = \bar{\mu} + \tau_k + \varepsilon_{ki} \qquad (7-7)$$

والتي يمكن كتابتها على الشكل التالي:

(حد الخطأ العشوائي) + (تأثير المجتمع k) + (التوقع الكلي) = (القياس x_{ki})

حيث أن: $\varepsilon_{ki} = x_{ki} - \mu_k$ ، وهو حد الخطأ العشوائي للقياس x_{ki} (أو البواقي)، وهي حدود مستقلة ويفترض أن تكون خاضعة للتوزيع الطبيعي $N(0 , \sigma^2)$. ولكن التأثيرات المجتمعية τ_k المعرفة في العلاقة (7–7) مرتبطة مع بعضها البعض , وذلك لأنه اعتماداً على العلاقة (7–2) نجد أن:

$$\sum_{k=1}^{g} n_k \tau_k = \sum_{k=1}^{g} n_k(\mu_k - \bar{\mu}) = \sum n_k\mu_k - \bar{\mu}\sum n_k = \sum n_k\mu_k - n\mu = 0$$

وبالتالي نحصل على أن:

$$\sum_{k=1}^{g} n_k * \tau_k = 0 \qquad (8-7)$$

وهذا يعني أن مجموع قيم التأثيرات τ_k المثقلة بأحجام العينات n_k يساوي الصفر. وبالتالي يكون متوسطها المثقل $\bar{\tau} = 0$

وبناء على التركيب (7–7), فإن تحليل التباين (ANOVA) في العينات يستخدم نموذجاً مشابهاً لـ (7–7)، للتعبير عن القياسات x_{ki} المشاهدة في العينات المسحوبة من تلك المجتمعات وذلك كما يلي:

$$x_{ki} = \bar{X} + (\bar{X}_k - \bar{X}) + (x_{ki} - \bar{X}_k) = \bar{X} + \tilde{\tau}_k + e_{ki} \qquad (9-6)$$

والذي يمكن كتابته في العينة على الشكل التالي:

(البواقي) + (تقدير تأثير المجتمع k**) + (المتوسط الكلي للعينة) = (القياس x_{Ki} المشاهد)** (10–7)

حيث أن: $\bar{X}$ هو المتوسط الكلي في العينة ويحسب من العلاقة: $\bar{X} = \frac{1}{n}\sum n_k\bar{X}_k$ ، ويعتبر $\bar{X}$ تقديراً غير متحيز للتوقع الكلي في المجتمع $\bar{\mu}$.

وأن: $\tilde{\tau}_k = (\bar{X}_k - \bar{X})$ هو تقدير لحد التأثير τ_k للمجتمع k، علماً بأن هذه الحدود يجب أن تحقق الشرط (7–8) التالي: $\sum n_k\tau_k = 0$.

وأن: $e_{ki} = (\bar{x}_{ki} - \bar{X}_k)$ هو تقدير لحد الخطأ العشوائي ε_{ki} ، ويسمى بحد البواقي(residual) في العينة k ، وهذه الحدود تشكل متحولات عشوائية, توقعاتها تساوي الصفر وتخضع لـ $N(0\ ,\sigma^2)$.

مثال: (7–1): لنفترض أننا نريد دراسة تغيرات أحد المتحولات X في 3 مجتمعات, فسحبنا منها ثلاث عينات بحجوم مختلفة هي: $n_1 = 3\ ,\ n_2 = 2\ ,\ n_3 = 3$. وبعد أخذ قياسات X من عناصر هذه العينات حصلنا على القياسات التالية:

$(n_1 = 3)$: المجتمع 1 : X_{1i} 9 , 6 , 9

$(n_2 = 2)$: المجتمع 2 : X_{2i} 0 , 2 ,

$(n_3 = 3)$: المجتمع 3 : X_{3i} 3 , 1 , 2

وعند حساب متوسطات X في هذه العينات نجد أنها تساوي ما يلي:

$$\bar{X}_1 = \frac{9+6+9}{3} = 8$$
$$\bar{X}_2 = \frac{0+2}{2} = 1$$
$$\bar{X}_3 = \frac{3+1+2}{3} = 2$$

وكذلك نجد أن المتوسط الكلي لها يساوي:

$$\bar{X} = \frac{1}{n}\sum n_k\bar{X}_k = \frac{3*8+2*1+3*2}{3+2+3} = 4$$

وللتحقق من صحة العلاقة (6–9) نحسب قيمتي القياسين x_{11} وx_{31} فنجد أن:

$$9 = x_{11} = \bar{X} + (\bar{X}_1 - \bar{X}) + (x_{11} - \bar{X}_1) = 4 + (8-4) + (9-8) = 4+4+1 = 9$$

....

$$3 = x_{31} = \bar{X} + (\bar{X}_3 - \bar{X}) + (x_{31} - \bar{X}_3) = 4 + (2-4) + (3-2) = 4-2+1 = 3$$

وهكذا يتم حساب بقية القيم x_{ki} .

وإذا قمنا بتطبيق مثل تلك الحسابات على كل مشاهدة من المشاهدات السابقة، نحصل على التصفيفة (arrays) التالية (وليس المصفوفة):

$$\begin{pmatrix} 9 & 6 & 9 \\ 0 & 2 & - \\ 3 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 4 & 4 \\ 4 & 4 & - \\ 4 & 4 & 4 \end{pmatrix} + \begin{pmatrix} 4 & 4 & 4 \\ -3 & -3 & - \\ -2 & -2 & -2 \end{pmatrix} + \begin{pmatrix} 1 & -2 & 1 \\ -1 & 1 & - \\ 1 & -1 & 0 \end{pmatrix}$$

(تصفيفة المشاهدات) = (تصفيفة المتوسط الكلي) + (تأثير تصفيفة المجتمعات (المعالجات)) + (تصفيفة البواقي)

لأن : $(x_{ki}) = (\bar{X}) + (\bar{X}_K - \bar{X}) + (x_{Ki} - \bar{X}_K)$

وبذلك نجد أن السؤال عن تساوي المتوسطات يتحول إلى تحديد فيما إذا كانت مساهمة تصفيفة تأثير المجتمعات (المعالجات) أكبر نسبياً من تصفيفة البواقي، علماً بأن التقديرات $\tilde{\tau}_k = (\bar{X}_k - \bar{X})$ للتأثيرات τ_k يجب أن تحقق دائماً الشرط (6−8) التالي:

$$\sum n_k \tilde{\tau}_k = 0$$

وضمن فرضية العدم H_0 ، يكون كل من $\tilde{\tau}_k$ تقديراً للعدد صفر، وهذا يعني أن تأثيرات المجتمعات تكون صغيرة.

أما إذا كانت تأثيرات المجتمعات (تأثيرات المعالجات) كبيرة، فإن ذلك سيؤدي إلى رفض الفرضية H_0 ، ولقياس مقدار مساهمة كل تصفيفة نكتب سطورها على شكل شعاع واحد، ثم نحسب مربع طول ذلك الشعاع، ويسمى هذا المقدار الجديد بمجموع المربعات (Sum of Squares) ونرمز له بالرمز SS .

فمثلاً نجد بالنسبة لتصفيفة القياسات المشاهدة (Observations) أن منقول شعاع عناصرها يكتب كما يلي:

$$Y' = [9\ \ 6\ \ 9\ \ 0\ \ 2\ \ 3\ \ 1\ 2]$$

وهو شعاع في الفضاء R^8 ، وإن مربع طوله يساوي $||Y||^2$ ويحسب كما يلي:

$$||Y||^2 = SS_{obs} = 9^2 + 6^2 + 9^2 + 0^2 + 2^2 + 3^2 + 1^2 + 2^2 = 216$$

وكذلك نجد أن مربع طول شعاع تصفيفة المتوسط الكلي يساوي:

$$SS_{means} = 4^2 + 4^2 + 4^2 + 4^2 + 4^2 + 4^2 + 4^2 + 4^2 = 128$$

وإن مربع طول شعاع تصفيفة تأثيرات المجتمعات أو (المعالجات treatments) يساوي :

$$SS_{tr} = 4^2 + 4^2 + 4^2 + (-3)^2 + (-3)^2 + (-2)^2 + (-2)^2 + (-2)^2 = 78$$

وإن مربع طول شعاع تصفيفة البواقي (residual) يساوي:

$$SS_{res} = 1^2 + (-2)^2 + 1^2 + (-1)^2 + 1^2 + 1^2 + (-1)^2 + 0^2 = 10$$

وبذلك نجد أن هذه المجاميع للمربعات ترتبط بنفس التركيب المعرف في (6−9) للمشاهدات، وهي التي تساوي:

$$SS_{obs} = SS_{means} + SS_{tr} + SS_{res}$$

حيث نلاحظ أن قيمها العددية تساوي:

$$216 = 128 + 78 + 10$$

أي أن تغيرات X الناتجة عن هذه المجتمعات تتوزع إلى ثلاث مركبات هي: تأثيرات المتوسطات والمعالجات والبواقي.

وإن تحليل التباين يقوم على مقارنة المقدار SS_{tr} (للمعالجات) مع المقدار SS_{res} (للبواقي)، فإذا كانت SS_{tr} أكبر بكثير من SS_{res} نرفض الفرضية H_0 ونقبل H_1 ونعتبر أن تغيرات X في هذه المجتمعات متباينة أو مختلفة.

والآن سنقوم باستخراج المعادلات الرياضية لهذه العلاقات وننطلق من العلاقة (7−9) التالية:

$$x_{ki} = \bar{X} + (\bar{X}_k - \bar{X}) + (x_{ki} - \bar{X}_k) \qquad (10-7)$$

ثم نطرح من الطرفين المتوسط الكلي $\bar{X}$ فنحصل على الانحرافات المصححة (أو الممعيرة) التالية :

$$(x_{ki} - \bar{X}) = 0 + (\bar{X}_k - \bar{X}) + (x_{ki} - \bar{X}_k)$$

وبذلك تختفي تصفيفة المتوسطات، ثم نقوم بتربيع الطرفين فنحصل على أن:

$$(x_{ki} - \bar{X})^2 = (\bar{X}_k - \bar{X})^2 + (x_{ki} - \bar{X}_k)^2 + 2(\bar{X}_k - \bar{X})(x_{ki} - \bar{X}_k) \qquad (11-7)$$

ثم نأخذ مجموع الحدود في الطرفين، المأخوذة على عناصر كل عينة k (أي $\sum_{i=1}^{n_k}$)، فنجد أنه ضمن كل عينة k يكون لدينا ما يلي:

$$\sum_{i=1}^{n_k}(x_{ki} - \bar{X})^2 = \sum_{i=1}^{n_k}(\bar{X}_k - \bar{X})^2 + \sum_{i=1}^{n_k}(x_{ki} - \bar{X}_k)^2 + 2(\bar{X}_k - \bar{X})\sum_{i=1}^{n_k}(x_{ki} - \bar{X}_k) \qquad (12-7)$$

وهنا نلاحظ أن المجموع الأول في الطرف الأيمن ليس له علاقة بدليل القياسات i , لذلك فهو يساوي:

$$\sum_{i=1}^{n_k}(\bar{X}_k - \bar{X})^2 = n_k(\bar{X}_k - \bar{X})^2$$

كما نلاحظ أن المجموع الأخير يساوي الصفر: $\sum_{i=1}^{n_k}(\mathrm{x}_{ki} - \bar{X}_k) = 0$

لأنه يمثل مجموع انحرافات قياسات X في العينة k عن متوسطها $\bar{X}_k$. وهذا يؤدي إلى انعدام الحد الأخير بكامله.

وبذلك تأخذ العلاقة (7−12) الشكل التالي:

$$\sum_{i=1}^{n_k}(x_{ki} - \bar{X})^2 = n_k(\bar{X}_k - \bar{X})^2 + \sum_{i=1}^{n_k}(x_{ki} - \bar{X}_k)^2 + 0 \qquad (13-7)$$

وذلك ضمن كل عينة k مسحوبة من ذلك المجتمع.

ثم نقوم بأخذ مجموع الحدود في الطرفين، المأخوذ على جميع المجتمعات (أي $\sum_{k=1}^{g}$) فنجد أن:

$$\sum_{k=1}^{g}\sum_{i=1}^{n_k}(x_{ki} - \bar{X})^2 = \sum_{k=1}^{g} n_k(\bar{X}_k - \bar{X})^2 + \sum_{k=1}^{g}\sum_{i=1}^{n_k}(x_{ki} - \bar{X}_k)^2 \qquad (14-7)$$

وهذا يعني أن هذه الأطراف حسب المفاهيم السابقة تساوي ما يلي:

(اجمالي مجموع المربعات المصحح بعد طرح $\bar{X}$) = (مجموع المربعات بين العينات المصحح) + (مجموع مربعات البواقي داخل العينات)

ونرمز لهذه المجاميع بالرموز التالية :

$$SST = SSB + SSW \qquad (15-7)$$

وهنا نشير إلى أن درجة حرية الحد SSB تساوي $(g-1)$، لأن المتوسطات $\bar{X}_k$ ترتبط مع المتوسط العام $\bar{X}$ بالعلاقة $\left(\bar{X} = \frac{1}{n}\sum n_k\,\bar{X}_k\right)$، وهذا ما ينقص عدد درجات الحرية بمقدار واحد وتصبح $v_1 = (g-1)$، ولإيجاد درجة حرية الحد الأخير SSW، نأخذ المجموع الداخلي منه $\sum_{i=1}^{n_k}(x_{ki} - \bar{X}_k)^2$ فنجد أن درجة حريته تساوي $(n_k - 1)$، لأن قياسات كل عينة x_{ki} مرتبطة بمتوسطها $\bar{X}_k$ وفق العلاقة

$\left(\bar{X}_k = \frac{1}{n_k}\sum x_{ki}\right)$، ومن ذلك نجد أن درجة حرية المجموع المضاعف $\sum_{k=1}^{g}\left[\sum_{i=1}^{n_k}(x_{ki}-\bar{X}_k)^2\right]$ تساوي مجموع درجات الحرية لما بداخله ، أي أنها تساوي:

$$v_2 = \sum_{k=1}^{g}(n_k - 1) = \sum_{k=1}^{g} n_k - g = n - g \qquad (16-7)$$

ولحساب درجة حرية الطرف الأيسر SST ، نأخذ مجموع درجات الحرية الطرف الأيمن، وبذلك تكون درجة حرية SST مساوية لما يلي: $v = v_1 + v_2$ أي أن:

$$v = (g-1) + (n-g) = n - 1 \qquad (17-7)$$

وأخيراً نضع نتائج هذه الحسابات في جدول منظم كالتالي:

جدول (7−2): الجدول النموذجي لـ ANOVA باتجاه واحد One-way

مصدر التباين	مجموع مربعات الانحرافات	درجة الحرية	متوسط مجموع المربعات	قيمة F
بين العينات (للمعالجات)	$SSB = \sum_{k=1}^{g} n_k(\bar{X}_k - \bar{X})^2$	$v_1 = g - 1$	$M\ SSB = \frac{SSB}{(g-1)}$	$F = \frac{M\ SSB}{M\ SSW}$
داخل العينات (الخطأ)	$SSW = \sum_{k}\sum_{i}(x_{ki} - \bar{X}_k)^2$	$v_2 = \sum_{k=1}^{g} n_k - g$	$M\ SSW = \frac{SSW}{(\sum n_k - g)}$	
المجموع الكلي المصحح	$SST = \sum_{k}^{g}\sum_{i}^{n_k}(x_{ki} - \bar{X})^2$	$v = \sum_{k=1}^{g} n_k - 1$		

ولاختبار صحة الفرضية H_0 نستخدم المؤشر F المعرف بالعلاقة التالية:

$$F = \frac{SSB/(g-1)}{SSW/(\sum n_k - g)} = \frac{M\ SSB}{M\ SSW} \qquad (18-7)$$

ثم نقارن هذه القيمة المحسوبة مع القيمة الحرجة $F_{v_1 v_2}(\propto)$ لمتحول التوزيع F المقابلة لمستوى الدلالة $\propto$ ولدرجتي الحرية:

$$v_1 = g - 1 \qquad\qquad v_2 = \left(\sum_{k=1}^{g} n_k - g\right) = n - g$$

ونتخذ القرار كما يلي:

(7 − 19) إذا كانت $F \le F_{v_1 v_2}(\propto)$ نقبل H_0 باحتمال ثقة قدره $(1 - \propto)$.

(7 − 20) أما إذا كانت $F > F_{v_1 v_2}(\propto)$ نرفض H_0 ونقبل H_1 بمستوى دلالة قدره $\propto$.

مثال (7−2): لنأخذ بيانات المثال (6−1) السابق والتي تتناول قياسات متحول واحد X من (3) عينات مسحوبة من (3) مجتمعات والتي كانت كما يلي:

$$\begin{matrix} P_1: (n_1 = 3) \\ P_2: (n_2 = 2): \\ P_3: (n_3 = 3) \end{matrix} \begin{pmatrix} 9 & 6 & 9 \\ 0 & 2 & - \\ 3 & 1 & 2 \end{pmatrix} : \begin{matrix} \bar{x}_1 = 8 \\ \bar{x}_2 = 1 \\ \bar{x}_3 = 2 \end{matrix} \quad , \quad \bar{X} = 4$$

ومنها نجد أن :

$$SSB = n_1(\bar{X}_1 - \bar{X})^2 + n_2(\bar{X}_2 - \bar{X})^2 + n_3(\bar{X}_3 - \bar{X})^2$$
$$= 3(8-4)^2 + 2(1-4)^2 + 3(2-4)^2 = 78$$

$$SSW = \sum_{k=1}^{g} \sum_{i=1}^{n_g} (x_{ki} - \bar{X}_k)^2 = [(9-8)^2 + (6-8)^2 + (9-8)^2] +$$
$$+[(0-1)^2 + (2-1)^2] + [(3-2)^2 + (1-2)^2 + (2-2)^2] = 10$$

وبذلك نجد أن قيمة SST (المصححة) تساوي:

$$SST = 78 + 10 = 88$$

ثم ننظم جدولاً خاصاً بذلك كما يلي :

جدول (7−3): تحليل التباين ANOVA باتجاه واحد One-way .

مصدر التباين	مجموع مربعات الانحرافات	درجة الحرية	متوسط مجموع المربعات	قيمة F
بين العينات (للمعالجات)	$SSB = 78$	$g - 1 = 2$	$M\ SSB = \frac{78}{2} = 39$	$F = \frac{M\ SSB}{M\ SSW} = \frac{39}{2}$
داخل العينات (الخطأ)	$SSW = 10$	$\sum n_k - g = 5$	$M\ SSW = \frac{10}{5} = 2$	
التباين الكلي (المصحح)	$SST = 88$	$\sum n_k - 1 = 7$		

ومنه نجد أن قيمة F المحسوبة تساوي:

$$F = \frac{M\ SSB}{M\ SSW} = \frac{\frac{SSB}{g-1}}{\frac{SSW}{\sum n_k - g}} = \frac{\frac{78}{2}}{\frac{10}{5}} = \frac{39}{2} = 19.5$$

ومن جداول التوزيع F نجد أن القيمة الحرجة لـ $F_{v_1 v_2}(\propto)$ عندما تكون $\propto = 0,05$ والمقابلة لدرجتي الحرية $v_1 = 2$ و$v_2 = 5$ ، تساوي: $F_{2,5}(0,05) = 5.786$: وبمقارنة F المحسوبة مع $F_{2,5}(\propto)$ الحرجة نجد أن:

$$(F = 19,5) > \left(F_{2,5}(0,05) = 5.786\right)$$

لذلك نرفض فرضية العدم H_0 التي تقول أن: $H_0 : \tau_1 = \tau_2 = \tau_3 = 0$ بمستوى دلالة 0.05 ، ونقبل H_1 التي تقول بوجود فروقات بين متوسطات تلك المجتمعات، وهنا يجب أن نتابع البحث عن مصدر تلك الفروقات.

ملاحظة 1: مما سبق نستنتج أنه يتم رفض H_0 عندما تأخذ النسبة $\frac{M\ SSB}{M\ SSW}$ قيمة كبيرة (أكبر من $F(\propto)$)، أو عندما تكون قيمة المقدار $\left(1 + \frac{M\ SSB}{M\ SSW}\right)$ كبيرة أيضاً، وهذا يكافئ القول التالي: إننا نرفض H_0 عندما تكون قيمة مقلوب المقدار السابق صغيرة . أي أننا نرفض H_0 عندما تكون قيمة المقدار التالي:

$$\lambda = \frac{1}{1 + \frac{M\,SSB}{M\,SSW}} = \frac{M\,SSW}{M\,SSB + M\,SSW} \qquad (20\,a - 7)$$

صغيرة بقدر كاف لتحقيق مستوى الدلالة $\propto$.

ويستفاد من هذه العلاقة في إيجاد العلاقة , التي سوف نستخدمها في رفض H_0 في حالة عدة متحولات, كما سنرى لاحقاً.

ملاحظة 2: حول علاقة مؤشر تحليل التباين F باختبار (ستودينت) t : عندما يكون لدينا مجتمعان فقط $(g = 2)$, فإنه يمكننا كتابة الحد SSB كما يلي:

$$SSB = n_1(\bar{X}_1 - \bar{X})^2 + n_2(\bar{X}_2 - \bar{X})^2 \qquad (20\,b - 7)$$

علماً بأن المتوسط الكلي $\bar{X}$ يحسب من العلاقة التالية:

$$\bar{X} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2}$$

وبتعويض $\bar{X}$ في $(20\,b - 7)$ نحصل على أن :

$$SSB = n_1\left(\bar{X}_1 - \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2}\right)^2 + n_2\left(\bar{X}_2 - \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2}\right)^2$$

$$SSB = \frac{n_1[(n_1 + \mathrm{n}_2)\bar{X}_1 - (n_1\bar{X}_1 + n_2\bar{X}_2)]^2 + n_2[(n_1 + n_2)\bar{X}_2 - (n_1\bar{X}_1 + n_2\bar{X}_2)]^2}{(n_1 + n_2)^2}$$

$$SSB = \frac{n_1\;n_2(\bar{X}_1 - \bar{X}_2)^2}{n_1 + n_2} = \frac{(\bar{X}_1 - \bar{X}_2)^2}{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

وبذلك نجد أن المؤشر F (عندما $g = 2$) يأخذ الشكل التالي:

$$F = \frac{\frac{SSB}{g - 1}}{\frac{SSE}{n - g}} = \frac{\frac{SSB}{1}}{\frac{SSE}{n_1 + n_2 - 2}} = \frac{(\bar{X}_1 - \bar{X}_2)^2}{\frac{SSE_1 + SSE_2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

وبما أن $SSE_1 = (n_1 - 1)S_1^2$ و $SSE_2 = (n_2 - 1)S_2^2$ نجد أن:

$$F_{1\,,n-2}(\propto) = \frac{(\bar{X}_1 - \bar{X}_2)^2}{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = t_{n-2}^2\left(\frac{\propto}{2}\right): \qquad (20\,C - 7)$$

وبمقارنة العبارة الأخيرة لـ F مع المؤشر t نجد: $F_{1\,,n-2}(\propto) = t_{n-2}^2\left(\frac{\propto}{2}\right)$ ، وهذا يعني أن مؤشر تحليل التباين F لمجتمعين يكافئ مربع اختبار (ستودينت) t، كما يمكن اعتبار الاختبار F تعميماً لاختبار (ستودينت) t .

2-7: تحليل التباين البسيط باتجاهين (n مشاهدة لكل خلية) (two-way ANOVA):

إن تحليل التباين البسيط باتجاهين يختلف جوهرياً عن تحليل التباين البسيط باتجاه واحد، وهو يستخدم لدراسة تغيرات متحول واحد (X) الناتجة عن تأثير عاملين نوعيين F_A و F_B (وليس عن مجتمعين كما في حالة الاتجاه الواحد)، وإن كل من هذين العاملين يأخذ عدة حالات، يدخلهما أو يتحكم بهما الباحث خلال مجربات التجارب، ثم يدرس تأثيرهما على المتحول التابع X. وهكذا يتم تنفيذ معظم الأبحاث العلمية.

فمثلاً: يمكن للباحث أن يدرس تغيرات أسعار طراز معين من الفساتين بتأثير عاملين: نوع القماش F_A وشكله الفني F_B ، كما يمكنه أن يدرس تغيرات الجاذبية الأرضية بتأثير درجة الطول F_A ودرجة العرض F_B .

وفي هذه الحالة يتوجب على الباحث تحديد الحالات أو القيم التي يأخذها كل من F_A و F_B، وأن يرسم جدولاً خاصاً لتقاطعاتهما، ثم عليه أن يجري تجربة واحدة على الأقل مقابل كل حجرة لتقاطعهما، ثم عليه وضع نتائج تلك التجارب في جدول مناسب لحالات تقاطع F_A و F_B .

ولنفترض الآن أن F_A يأخذ g حالة منفصلة، وأن F_B يأخذ q حالة منفصلة، وان الباحث قد أجرى تجربة واحدة مقابل كل تقاطع لهما ووضع نتائجه في جدول كالتالي:

جدول (7-4): الحالات المتقاطعة لـ F_A و F_B

F_A / F_B	1	2	3	⋯	g
1	x_{11}	x_{21}	x_{31}	⋯	x_{g1}
2	x_{12}	x_{22}	x_{32}	⋯	x_{g1}
⋮	⋮	⋮	⋮	⋯	⋮
q	x_{1q}	x_{2q}	x_{3q}	⋯	x_{gq}

ولكن الباحث يمكن أن يقوم بتكرار تجاربه مقابل كل حجرة (k,ℓ) عدداً من المرات، وليكن n مرة، فعندها سيحصل في كل حجرة على n نتيجة . وهذه النتائج تمثل عينة من القياسات مأخوذة من مجتمع التجارب الذي يقابل كل حجرة (k,ℓ) . وبذلك يكون لدينا $(g*q)$ مجتمعاً إحصائياً سحبت منها $(g*q)$ عينة عشوائية بحجوم متساوية n قياساً لـ X في كل منها.

ولنفترض أن توقع X في كل حجرة (k,ℓ) منها يساوي $(\mu_{k\ell})$, وأن التوقع الكلي لـ X لها يساوي μ ، حيث أن:

$$\mu = \frac{\sum_{k=1}^{g}\sum_{\ell=}^{q}\mu_{k\ell}}{g*q} \qquad (21-7)$$

وعندها يمكننا أن نعبر عن التوقع $\mu_{k\ell}$ في الحجرة (k,ℓ) بعلاقة مركبة كما يلي:

$$\mu_{k\ell} = \mu + (\mu_{k\ell} - \mu) \qquad (22-7)$$

وحتى نظهر تأثير العاملين F_A و F_B وتأثيرهما المشترك $(F_A\ F_B)$ على نتائج تلك التجارب نكتب التوقع $\mu_{K\ell}$ على الشكل التالي (وذلك بإضافة وطرح التوقعات الهامشية μ_k و μ_ℓ من الطرف الأيمن):

$$\mu_{k\ell} = \mu + (\mu_k - \mu) + (\mu_\ell - \mu) + (\mu_{k\ell} - \mu_k - \mu_\ell + \mu) \qquad (23-7)$$

وهنا نلاحظ أن الأقواس تعكس تأثير العامل F_A والعامل F_B وتأثيرهما معاً، واختصاراً للرموز نكتب ذلك كما يلي:

$$\mu_{k\ell} = \mu + \propto_k + \beta_\ell + \gamma_{k\ell} \qquad (24-7)$$

حيث أن: $\propto_k = (\mu_k - \mu)$ و $\beta_\ell = (\mu_\ell - \mu)$ و $\gamma_{k\ell} = (\mu_{k\ell} - \mu_k - \mu_\ell + \mu)$

وهنا يشترط على المقادير $\propto_k$ و β_ℓ و $\gamma_{k\ell}$ أن تحقق الشروط التالية (يمكن البرهان على ذلك كما فعلنا في (7–8) مع ملاحظة أن حجوم العينات هنا متساوية وتساوي n).

$$\sum_{k=1}^{g} \propto_k = 0 \qquad \sum_{\ell=1}^{q} \beta_\ell = 0$$
$$\sum_{k=1}^{g} \gamma_{k\ell} = 0 \qquad \sum_{\ell=1}^{q} \gamma_{k\ell} = 0 \qquad (25-7)$$

والآن نعود إلى العلاقة (7–24) ونكتبها كما يلي:

$$\mu_{k\ell} = E(x_{k\ell i}) = \mu + \propto_\text{k} + \beta_\ell + \gamma_{k\ell} \qquad (26-7)$$

وفي العينات يمكننا كتابة قيمة أي قياس $x_{K\ell i}$ بخطأ $e_{K\ell i}$ كما يلي:

$$x_{k\ell i} = \bar{X} + \propto_k + \beta_\ell + \gamma_{k\ell} + e_{k\ell i} \qquad (27-7)$$

أو كما يلي :

$$x_{K\ell i} = \bar{X} + (\bar{X}_k - \bar{X}) + (\bar{X}_\ell - \bar{X}) + (\bar{X}_{K\ell} - \bar{X}_k - \bar{X}_\ell - \bar{X}) + (x_{K\ell i} - \bar{X}_{K\ell}) \qquad \mathbf{(28-7)}$$

$$\begin{pmatrix}\text{قيمة القياس}\\ x_{K\ell i}\end{pmatrix} = \begin{pmatrix}\text{المتوسط}\\ \text{الكلي}\end{pmatrix} + \begin{pmatrix}\text{تأثير العامل}\\ \text{الأول } F_A\end{pmatrix} + \begin{pmatrix}\text{تأثير العامل}\\ \text{الثاني } F_B\end{pmatrix} + \begin{pmatrix}\text{تأثير تداخل العاملين}\\ F_A \text{ و } F_B \text{ معا}\end{pmatrix} + \begin{pmatrix}\text{الخطأ العشوائي}\\ \text{أو البواقي}\end{pmatrix}$$

حيث أن: $\bar{X}_k$ ترمز لمتوسطات العامل الأول F_A وأن $k : 1\ 2\ 3 \ldots g$

وأن: $\bar{X}_\ell$ ترمز لمتوسطات العامل الثاني F_B وأن $\ell : 1\ 2\ 3 \ldots q$

وأن: $\bar{X}_{k\ell}$ ترمز لمتوسطات القياسات في الحجرة (k, ℓ): $\bar{X}_{k\ell} = \frac{1}{n}\sum_{i=1}^{n} x_{k\ell i}$

وأن: $e_{k\ell i}$ هي قيم البواقي أو الخطأ العشوائي، وهي عبارة عن متحولات عشوائية مستقلة ضمن كل حجرة وخاضعة للتوزيع الطبيعي $N(0\,, \sigma^2)$ الذي توقعه يساوي الصفر ولها تباين موحد يساوي σ^2 .

وعندها فإن الفرضيات البحثية تأخذ الشكل التالي: فرضية العدم: وهي تتألف مما يلي:

$$H_0 : \begin{cases} \propto_k = 0 & k : 1\ 2\ 3 \ldots g \\ \beta_\ell = 0 & \ell : 1\ 2\ 3 \ldots q \\ \gamma_{k\ell} = 0 & k, \ell : 1\ 2\ 3 \ldots \end{cases} \qquad (29-7)$$

أما الفرضية البديلة فتكون كما يلي:

$$H_1 : \begin{cases} \propto_K \neq 0 & \text{من أجل } k \text{ واحدة على الأقل :} \\ \beta_\ell \neq 0 & \text{من أجل } \ell \text{ واحدة على الأقل :} \\ \gamma_{k\ell} \neq 0 & \text{من أجل زوج } (\ell\, , k) \text{ واحد على الأقل :} \end{cases} \qquad (30-7)$$

ولاستخراج مؤشرات الاختبار المناسبة نقوم بمعالجة العلاقة السابقة (7–28) كما فعلنا مع العلاقة السابقة (7–12) فنحصل على العلاقة التالية:

$$\sum_{k=1}^{g}\sum_{\ell=1}^{q}\sum_{i=1}^{n}(x_{k\ell i}-\bar{X})^2 = qn\sum_{k=1}^{g}(\bar{X}_k-\bar{X})^2 + gn\sum_{\ell=1}^{q}(\bar{X}_\ell-\bar{X})^2 +$$

$$+n\sum_{k=1}^{g}\sum_{\ell=1}^{q}(\bar{X}_{k\ell}-\bar{X}_k-\bar{X}_\ell+\bar{X})^2 + \sum_{k=1}^{g}\sum_{\ell=1}^{q}\sum_{i=1}^{n}(x_{k\ell i}-\bar{X}_{k\ell})^2 \qquad (31-7)$$

والتي سنرمز لأطرافها اختصاراً كما يلي:

$$SST = SSA + SSB + SSAB + SSE \qquad (32-7)$$

أما بالنسبة لدرجات الحرية التي تقابل كل منها فهي تساوي:

$$(g*q*n-1) = (g-1)+(q-1)+(g-1)(q-1)+gq(n-1) \qquad (33-7)$$

ثم نعرف مؤشرات الاختبار المناسبة لكل من F_A و F_B وللتداخل بينهما كما يلي:

$$F_A = \frac{\frac{SSA}{g-1}}{\frac{SSE}{gq(n-1)}} = \frac{gq(n-1)SSA}{(g-1)SSE} \qquad \left(\text{للعامل } F_A\right) \qquad (\mathbf{34-7})$$

وهو يخضع للتوزيع F بدرجتي حرية $v_1 = (g-1)$ و $v_2 = g*q(n-1)$.

$$F_B = \frac{\frac{SSB}{q-1}}{\frac{SSE}{gq(n-1)}} = \frac{gq(n-1)SSB}{(q-1)SSE} \qquad \left(\text{للعامل } F_B\right) \qquad (\mathbf{35-7})$$

وهو يخضع للتوزيع F بدرجتي حرية: $v_1 = (q-1)$ و $v_2 = g*q(n-1)$.

أما مؤشر اختبار التداخل بين F_A و F_B فيعرف كما يلي:

$$F_{AB} = \frac{\frac{SSAB}{(g-1)(q-1)}}{\frac{SSE}{gq(n-1)}} = \frac{gq(n-1)SSAB}{(g-1)(q-1)SSE} \qquad (\mathbf{36-7})$$

وهو يخضع للتوزيع F بدرجتي حرية $v_1 = (g-1)(q-1)$ و $v_2 = g*q(n-1)$.

ثم نقوم بتنظيم جدول مناسب لتحليل التباين البسيط باتجاهين كما يلي:

جدول (7–5): جدول ANOVA باتجاهين two way:

مصدر التباين	مجموع المربعات	درجة الحرية	المؤشر F
العامل F_A	$SSA =$ (عدد)	$g-1$	$F_A = \frac{gq(n-1)SSA}{(g-1)SSE}$
العامل F_B	$SSB =$ (عدد)	$q-1$	$F_B = \frac{gq(n-1)SSB}{(q-1)SSE}$
التداخل F_AF_B	$SSAB =$ (عدد)	$(g-1)(q-1)$	$F_{AB} = \frac{gq(n-1)SSAB}{(g-1)(q-1)SSE}$
البواقي أو الخطأ العشوائي	$SSE =$ (عدد)	$gq(n-1)$	———
الإجمالي (المصحح)	T= (عدد) SS	$gq\,n-1$	———

وبعدها نقوم بمقارنة كل من F_A و F_B و F_{AB} بالقيم الحرجة المقابلة لها $F_{v_1,v_2}(\propto)$ بدرجتي الحرية v_1 و v_2 و نتخذ القرار حسب العامل المفروض كما يلي :

إذا كانت $F \leq F_{v_1,v_2}(\propto)$ نقبل الفرضية H_0 حسب العامل المفروض

$(37-7)$

أما إذا كانت $F > F_{v_1,v_2}(\propto)$ نرفض الفرضية H_0 ونقبل H_1

ثم نستخلص النتائج الممكنة من هذه الاختبارات كما سنرى من خلال المثال التالي:

ملاحظة: إذا كانت نتيجة اختبار التداخل F_AF_B، هي رفض H_0 وقبول H_1، فإن ذلك يعني أن تداخل العاملين A,B ، يؤثر معنوياً على تغيرات المتحول المدروس X ، وفي هذه الحالة تكون عملية البحث عن تأثير كل منها بمفرده غير مجدية وتصبح غير واضحة وغير ممكنة التفسير، ومما سبق نستنتج أن تحليل التباين باتجاهين يكون مجدياً فقط عندما يكون تأثير التداخل غير معنوي.

مثال (7–3): لنفترض أننا نريد دراسة تغيرات أسعار طراز معين من الفساتين في السوق، ومعرفة درجة تأثر أسعارها X بعاملين هما: نوعية القماش F_A وشكل أوزخرفة الفستان F_B . وبعد الدراسة تبين لنا أن هذه الفساتين تُصنع من (3) أنواع من القماش هي $A(A_1\,, A_2\,, A_3)$، وأن شكلها يأخذ شكلين أساسين من الزخرفة هما $(B_1\,, B_2)$. ثم قمنا بتتبع أسعار هذه الفساتين حسب كل تقاطعات حالات النوع والشكل، ولذلك أخذنا (4) أسواق (محلات) تبيع هذه الفساتين وسجلنا الأسعار فيها حسب النوع والشكل, فكانت كما يلي (حسبنا متوسطات الاسعار في كل حجرة ووضعناها ضمن مستطيلات):

جدول (7–6): بيانات المثال (فرضية)

المتوسطات $X\ell$	الاجمالي		نوع ثالث A_3	نوع ثاني A_2	نوع أول A_1	أنواع القماش / أشكال الفساتين
447.5	5370		420 440 450 460 480	410 420 425 430 440	430 450 467.5 460 530	أحمر مزخرف B_1
392.5	4710		390 370 390 400 400	350 370 380 400 400	400 400 407.5 400 430	أبيض مزخرف B_2
	10080		3360	3220	3500	الاجمالي X_k
$\bar{X} = 420$			420	402.5	437.5	المتوسطات $\bar{X}_k$

والمطلوب دراسة تأثير العاملين A وB على السعر X . وإجراء الاختبارات اللازمة بمستوى دلالة ($\propto = 0.05$) ، علماً بأن الأرقام ضمن المستطيلات في كل حجرة هي متوسطات القياسات فيها $\bar{X}_{k\ell}$ وإننا حسبنا المتوسط الكلي لها فكان: $\bar{X} = 420$.

الحل: نقوم أولاً بحساب المجاميع التي في العلاقة (7–31) فنجد أن:

$$SST = \sum_{k=1}^{3}\sum_{\ell=1}^{2}\sum_{i=1}^{4}(x_{k\ell i} - 420)^2 = 34600$$

وذلك لأن:

$$SST = (100 + 900 + 1600 + 12100) + (100 + 0 - 100 + 400) + (0 + 400 + 1600 + 3600) + (400 + 400 + 400 + 100) + (4900 + 2500 + 400 + 400) + (900 + 2500 + 400 + 400) = 34600$$

ثم نقوم بحساب SSA من العلاقة:

$$SSA = qn\sum_{k=1}^{3}(\bar{X}_k - \bar{X})^2 =$$

$$SSA = 2 * 4[306.5 + 306.5 + 0] = 4900$$

ثم نقوم بحساب SSB من العلاقة:

$$SSB = gn\sum_{\ell=1}^{2}(\bar{X}_\ell - \bar{X})^2 =$$

$$SSB = 3 * 4[756.25 + 756.25] = 18150$$

ثم نقوم بحساب حد البواقي أو الخطأ العشوائي في جميع الحجر من العلاقة:

$$SSE = \sum_{k=1}^{3}\sum_{\ell=1}^{2}\sum_{i=1}^{4}(x_{k\ell i} - \bar{X}_{k\ell})^2 =$$

$$SSE = 5675 + 500 + 2000 + 675 + 1800 + 600 = 11250$$

وأخيراً نقوم بحساب حد التداخل SSAB من العلاقة:

$$SSAB = SST - SSA - SSB - SSE$$

$$SSAB = 34600 - 4900 - 18150 - 11250 = 300$$

ثم نقوم بتنظيم جدول التحليل التالي:

جدول (7−7): جدول تحليل ANOVA باتجاهين

مصدر التباين	مجموع المربعات	درجة الحرية	متوسطات المربعات	قيم F
نوعية القماش A	$SSA = 4900$	$g - 1 = 2$	$MSSA = 2450$	$FA = 3.92$
شكل الفستان B	$SSB = 18150$	$q - 1 = 1$	$MSSB = 18150$	$F_B = 29.04$
تداخل A وB	$SSAB = 300$	$(g-1)(q-1) = 2$	$MSSAB = 150$	$F_{AB} = 0.24$
الخطأ العشوائي	$SSE = 11250$	$gq(n-1) = 18$	$MSSE = 625$	———
الاجمالي	$SST = 34600$	$gqn - 1 = 19$	———	———

ولاختبار تأثير هذين العاملين وتأثير تداخلهما، علينا أن نقوم بإيجاد القيمة الحرجة لـ F المقابلة لكل حالة، فنجد أن:

$$F_{2,18}(0.05) = F_{2,18}(0.05) = 3.55 \quad \text{و} \quad F_{1,18}(0.05) = 4.41$$

ثم نقوم بمقارنة قيم F_A مع $F_{2,18}(\propto)$ فنجد أن: $(F_A = 3.92) > 3.55$, لذلك نرفض فرضية العدم H_0، التي تقول أنه لا يوجد تأثير لنوعية القماش على سعره , ونقبل H_1 التي تقول أن أسعار الفساتين تتأثر بنوعية القماش (بنسبة $\left(\frac{4900}{34600}\right) * 100\%$) .

ثم نقوم بمقارنة F_B مع $F_{1,18}(\propto)$ نجد أيضاً أن: $(F_B = 29.04) > 4.41$, , لذلك نرفض فرضية العدم التي تقول أنه لا يوجد تأثير لشكل الفستان على سعره , ونقبل H_1 التي تقول أن أسعار الفساتين تتأثر كثيراً بشكل القماش (بنسبة $\left(\frac{18150}{34600}\right) * 100\%$).

ولاختبار تأثير التفاعل الداخلي للعاملين A وB نقارن قيمة F_{AB} مع القيمة الحرجة $F_{2\,18}(\propto)$ فنجد أن:

$(F_{AB} = 0.24) < 3.55$, لذلك نقبل H_0 ونعتبر أن أسعار الفساتين لا تتأثر بتداخل العاملين المدروسين وهما النوعية والشكل.

ملاحظة: يفضل في التطبيقات العملية أن نبدأ بإجراء اختبار التداخل . فإذا كان تأثير ذلك التداخل معنوياً (حالة رفض H_0) . فإن ذلك يعني أن تداخل العاملين F_A وF_B يؤثر كثيراً على تغيرات المتحول التابع X . وهذا يجعل عملية اختبار تأثير كل من F_A وF_B على حدة على X غير واضحة وصعبة التفسير، وينصح بعدم متابعة دراسة تأثيراتهما المنفردة.

أما إذا كان تأثير التداخل مهملاً فإنه يمكننا متابعة التحليل وإجراء الاختبارين حول تأثير F_A وF_B واستخلاص النتائج الممكنة.

7−3: تحليل التباين البسيط بثلاث اتجاهات (n مشاهدة لكل خلية) :

إن تحليل التباين في هذه الحالة هو تعميم للحالة السابقة (7−2)، لذلك نفترض أننا نريد دراسة تغيرات أحد المتحولات X (المتحول التابع) الناتجة عن تأثير (3) عوامل نوعية: F_A وF_B وF_C ، وإن لكل من هذه العوامل عدة حالات يرمز لها كما يلي:

$$F_A : A_1, A_2, \dots, A_K, \dots, A_g$$
$$F_B : B_1, B_2, \dots, B_\ell, \dots, B_q \qquad (38-7)$$
$$F_C : C_1, C_2, \dots, C_m, \dots, C_r$$

ولذلك ننشأ مكعب التقاطعات الممكنة لهذه الحالات، فنحصل على $(g * q * r)$ حجرة، ثم نأخذ من كل حجرة منه n قياساً للمتحول المدروس X . فنحصل على $(g * q * r)$ عينة مسحوبة من مجتمعات تلك الحجر وعلى $(g * q * r * n)$ قياساً.

وإذا رمزنا لتلك القياسات بالرموز $x_{k\ell mi}$ فإنه يمكننا كتابة النموذج الرياضي الموافق لهذا التحليل باستخدام نفس المفاهيم والرموز السابقة كما يلي:

$$x_{k\ell mi} = \bar{X} + \propto_k + \beta_\ell + \gamma_m + (\propto\beta)_{k\ell} + (\propto\gamma)_{km} + (\beta\gamma)_{\ell m} + (\propto\beta\gamma)_{k\ell m} + (e_{k\ell mi}) \quad \mathbf{(39-7)}$$

ويشترط على هذه الرموز أن تحقق شروط مشابهة للشروط السابقة المفروضة على تحليل التباين باتجاهين والمذكورة في (7–25).

أما فرضية العدم H_0 (وهي عدم وجود تأثير لهذه العوامل على X) فتكتب كما يلي:

$$H_0 : \begin{cases} \propto_1 = \propto_2 = \cdots = \propto_g = 0 \\ \beta_1 = \beta_2 = \cdots = \beta_q = 0 \\ \gamma_1 = \gamma_2 = \cdots = \gamma_r = 0 \end{cases} \qquad (40-7)$$

أما الفرضية البديلة H_1 فتكتب كما يلي:

$$H_1 : \begin{cases} \propto_k \neq 0 & \text{من أجل } k \text{ واحدة على الأقل} \\ \beta_\ell \neq 0 & \text{من أجل } \ell \text{ واحدة على الأقل} \\ \gamma_m \neq 0 & \text{من أجل } m \text{ واحدة على الأقل} \end{cases} \qquad (41-7)$$

ثم نقوم بمعالجة العلاقة (7–39) كما فعلنا في الفقرات السابقة فنحصل على مجاميع المربعات المختلفة ونكتبها كما يلي:

$$SST = SSA + SSB + SSC + SS(AB) + SS(AC) + SS(BC) + SS(ABC) + SSE \quad (42-7)$$

وإن درجات الحرية المقابلة لهذه المجاميع هي كما يلي:

$$(gqrn) - 1 = (g-1) + (q-1) + (r-1) + (g-1)(q-1) + (g-1)(r-1)$$
$$+ (q-1)(r-1) + (g-1)(q-1)(r-1) + gqr(n-1) \qquad (43-7)$$

وسنقوم بتعريف هذه المجاميع وشرح كيفية حسابها لاحقاً من خلال المثال (7–4):

أما بالنسبة لإجراء الاختبارات اللازمة نبدأ بإجراء اختبار تأثير التداخل الثلاثي SS(ABC)، ونحسب قيمة مؤشر الاختبار F_{ABC} المقابل له من العلاقة:

$$F_{ABC} = \frac{\dfrac{SS(ABC)}{(g-1)(q-1)(r-1)}}{\dfrac{SSE}{gqr(n-1)}} = \frac{\dfrac{SS(ABC)}{v_1}}{\dfrac{SSE}{v_2}} \qquad (44-7)$$

ثم نقارن هذه القيمة المحسوبة مع القيمة الحرجة لمتحول $F_{v_1,v_2}(\propto)$ المقابل لمستوى الدلالة $\propto$ ولدرجتي الحرية v_1 و v_2، ونتخذ القرار كما يلي:

$$\text{إذا كانت } F \le F_{v_1,v_2}(\propto) \text{ نقبل فرضية العدم } H_0$$
$$\text{وإذا كانت } F > F_{v_1,v_2}(\propto) \text{ نرفض } H_0 \text{ ونقبل } H_1 \qquad (45-7)$$

وإذا كانت النتيجة رفض H_0 وقبول H_1، فإن ذلك يعني أن تداخل العوامل الثلاثة يؤثر معنوياً على تغيرات المتحول المدروس X ، وفي هذه الحالة تكون عملية البحث عن تأثير كل منها بمفرده غير مجدية وتصبح غير واضحة وغير ممكنة التفسير، ومما سبق نستنتج أن تحليل التباين بثلاث اتجاهات يكون مجدياً فقط عندما يكون تأثير التداخل غير معنوي.

أما إذا كانت نتيجة الاختبار السابق قبول H_0، فإن ذلك يعني أن تأثير تداخل العوامل الثلاثة غير معنوي ويمكن إهماله . وبعدها ننتقل إلى اختبارات تأثيرات التداخلات الثنائية . وإذا كانت غير معنوية، نقوم باختبار تأثيرات العوامل المنفردة وذلك بتطبيق نفس الإجراءات السابقة .

مثال (7–4): لدراسة مقاومة ألواح السيراميك للصدمات ثم تحديد (3) عوامل مؤثرة عليها وهي :

- نوع المادة الأولية وتأخذ حالتين فقط (A_1 , A_2) .
- مساحة اللوح ويأخذ حالتين أيضاً (B_1 , B_2) .
- سماكة اللوح ويأخذ (3) حالات هي : (C_1 , C_2 , C_3) .

ثم تم اختبار (5) ألواح من كل حجرة لتقاطعاتها وأجريت عليها تجارب لقياس المقاومة (بالكيلوغرام) فكانت نتائج تلك القياسات كما في الجدول التالي :

جدول (7–7): بيانات المثال (فرضية)

نوع المادة		نوع المادة A_1						نوع المادة A_2					
المساحة		المساحة B_1			المساحة B_2			المساحة B_1			المساحة B_2		
السماكة		C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3
عناصر العينة i	$i=1$	67	72	78	67	79	78	54	52	63	54	57	60
	2	66	67	81	71	80	78	51	56	54	56	58	68
	3	62	75	67	72	81	77	47	52	65	58	61	61
	4	71	70	75	70	80	83	51	52	62	51	59	61
	5	69	71	75	81	85	79	59	53	60	57	55	67

والمطلوب دراسة تأثير هذه العوامل على مقاومة ألواح السيراميك وإجراء الاختبارات اللازمة عليها بمستوى دلالة $\propto = 0.05$

الحل: لإجراء هذه الاختبارات نحتاج إلى كثير من الحسابات المعقدة لإيجاد قيم حدود العلاقة (6–42) ولذلك سنستخدم علاقات رياضية مبسطة (بدون برهانها)، وحتى نستطيع تطبيق تلك العلاقات أعددنا الجدولين (6–8) و (6–9) واستخدمنا بياناتهما في عملية الحساب كما يلي:

نقوم بحساب SST من العلاقة التالية :

$$SST = \sum_{k=1}^{g}\sum_{\ell=1}^{q}\sum_{m=1}^{r}\sum_{i=1}^{n}(x_{k\ell mi} - \bar{X})^2 = \sum_{k=1}^{g}\sum_{\ell=1}^{q}\sum_{m=1}^{r}\sum_{i=1}^{n} x_{k\ell mi}^2 - \frac{X^2}{g * q * r * n} \qquad (45-7)$$

ومن الجدولين (7–8) و(7–9) نجد أن:

$$SST = 265174 - \frac{(3942)^2}{2 * 2 * 3 * 5} = 265174 - 258989.4 = 6184.6$$

ومن السطر الأخير في الجدول (7–8) نجد مباشرة أن:

$$SSE = \sum_{k=1}^{g}\sum_{\ell=1}^{q}\sum_{m=1}^{r} SSE_{k\ell mi} = 614$$

ثم نقوم بحساب SSA من بيانات الجدول (7–9) وباستخدام العلاقة التالية:

$$SSA = \frac{1}{qrn}\sum_{k=1}^{2} X_k^2 - \frac{X^2}{gqrn} = \frac{1}{30}[(2228)^2 + (2044)^2] - 258989.4 = 4403.3 \qquad (46-7)$$

جدول (7–8): نتائج الحسابات:

النوع K	نوع المادة A_1						نوع المادة A_2						
المساحة ℓ	المساحة B_1			المساحة B_2			المساحة B_1			المساحة B_2			
السماكة m	C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3	المجموع
i=1	67	72	78	67	79	78	54	52	63	54	57	60	
2	66	67	81	71	80	78	51	56	54	56	58	68	
3	62	75	67	72	81	77	47	52	65	58	61	61	
4	71	70	75	70	80	83	51	52	62	51	59	61	
5	69	71	75	81	85	79	59	53	60	57	55	67	
$X_{k\ell m}$ المجموع	335	355	377	361	405	395	362	265	304	276	290	311	3942
$\frac{X_{k\ell m}^2}{n}$	22445	25205	28425.6	26064.2	32803	31205	13728.8	14045	18483.2	15235.2	16820	200988	
$\sum x_{k\ell m}^2$	22491	25239	26535	26173	32827	31227	13808	14057	18554	15266	16840	20155	$T = 265174$
$SSE_{k\ell m}$	46	36	109.2	1196	22	22	79.2	12	70.8	30.8	20	57.2	$SSE = 614$

الجدول (7−9): مجاميع المجاميع حسب الحالات السابقة:

المجاميع	$m = 1$	$m = 2$	$m = 3$	المجموع $X_{k\ell 1}$
X_{111}	335	355	377	= 1067
X_{121}	361	405	395	= 1161
X_{211}	362	265	304	= 831
X_{221}	276	290	317	= 883
$X_{k1} = \begin{cases} k = 1 \\ k = 2 \end{cases}$	696 538	760 555	772 621	= 2228 = 2044
$X_{\ell=2} = \begin{cases} \ell = 1 \\ \ell = 2 \end{cases}$	597 637	620 695	681 712	= 1898 = 2044
X_m	1234	1315	1393	$X = 3942$

ثم نقوم بحساب SSB بناءً على الجدول (6−9) وباستخدام العلاقة التالية:

$$SSB = \frac{1}{grn}\sum_{\ell=1}^{2} X_{\ell}^2 - \frac{X^2}{gqrn} = \frac{1}{30}[(1898)^2 + (2044)^2] - 258989.4 = 355.3 \qquad (47-7)$$

ثم نقوم بحساب SSC بناءً على الجدول (7−9) من العلاقة التالية:

$$SSC = \frac{1}{gqn}\sum X_m^2 - \frac{X^2}{gqrn} = \frac{1}{20}[(1234)^2 + (1315)^2 - (1393)^2] - 258989.4 = 632.1 \qquad (\mathbf{48-7})$$

ثم نقوم بحساب SS(AB) بناءً على العمود الأخير الجدول (7−9) من العلاقة التالية:

$$SS(AB) = \frac{1}{rn}\sum_{k=1}^{2}\sum_{\ell=1}^{2} X_{k\ell}^2 - \frac{X^2}{gqrn} - SSA - SSB \qquad (49-7)$$

$$= \frac{1}{15}[(1067)^2 + (1161)^2 + (831)^2 + (883)^2] - 258989.4 - 4403.3 - 355.3 = 29.3$$

ثم نقوم بحساب SS(AC) بناءً على الجدول (7−9) من العلاقة التالية:

$$SS(AC) = \frac{1}{qn}\sum_{k=1}^{2}\sum_{m=1}^{2} X_{km}^2 - \frac{X^2}{gqrn} - SSA - SSC \qquad (50-7)$$

$$= \frac{1}{10}[(696)^2 + (760)^2 + (772)^2 + (538)^2 + (555)^2 + (621)^2] - 258989.4 - 4403.3 - 632.1 = 68.2$$

ثم نقوم بحساب SS(BC) بناءً على الجدول (7−9) من العلاقة التالية:

$$SS(BC) = \frac{1}{gn}\sum_{\ell=1}^{2}\sum_{m=1}^{3} X_{\ell m}^2 - \frac{X^2}{gqrn} - SSB - SSC \qquad (51-7)$$

$$= \frac{1}{10}[(597)^2 + (620)^2 + (681)^2 + (637)^2 + (695)^2 + (712)^2] - 258989.4 - 355.3 - 632.1 = 54$$

وأخيراً نحسب حد التداخل الثلاثي SS(ABC) من العلاقة (7−42) فنجد أن:

$$SS(ABC) = 6184.6 - 614 - 4403.3 - 355.3 - 632.1 - 29.3 - 68.2 - 54 = 10.4$$

ثم نقوم بوضع نتائج هذه الحسابات في جدول تحليل التباين ذي الاتجاهات الثلاثة المرفقة بدرجات الحرية كما يلي:

جدول (7−10): جدول تحليل التباين بثلاث اتجاهات

مصدر التباين	رمز المصدر	درجة الحرية	مجموع المربعات	متوسط المربعات	$\tilde{F}$	$F(\propto)$
نوع المادة F_A	SSA	$g-1=1$	4403.3	4403.3	344	4.04
المساحة F_B	SSB	$q-1=1$	355.3	355.3	27.8	4.04
السماكة F_C	SSC	$r-1=2$	632.1	316.0	24.7	3.19
أثر المادة والمساحة	$SS(AB)$	$(g-1)(q-1)=1$	29.3	29.3	2.29	4.04
أثر المادة والسماكة	$SS(AC)$	$(g-1)(r-1)=2$	86.2	43.1	3.37	3.19
أثر المساحة والسماكة	$SS(BC)$	$(q-1)(r-1)=2$	54.0	27.0	2.11	3.19
أثر العوامل الثلاثة المادة والمساحة والسماكة	$SS(ABC)$	$(g-1)(q-1)(r-1)=2$	10.4	5.2	0.41	3.19
الخطأ العشوائي (البواقي)	SSE	$gqr(n-1)=48$	614.0	12.79	−	−
الإجمالي	SST	$(gqrn)-1=59$	6184.6	−	−	−

وبعد حساب قيم F لجميع هذه المجاميع نقارنها مع قيمة $F(\propto)$ الحرجة المقابلة لها, والمبينة في العمود الأخير من الجدول (7−10) فنلاحظ ما يلي:

1− إن تأثير التداخل الثلاثي غير معنوي لأن $F_{ABC}=0.41<3.19$.

2− إن تأثير التداخل الثنائي SS(AC) معنوي وهذا يؤثر على تفسير النتائج .

3− إن تأثير التداخلين الثنائيين(SS(AB و SS(BC) غير معنويين .

4− إن تأثير العوامل الثلاثة F_A وF_B وF_C معنوية وأن العامل F_A هو الأكثر تأثيراً على تغيرات المقاومة

7−4: تحليل المربع اللاتيني (بمشاهدة واحدة لكل خلية):

لقد لاحظنا في الفقرة السابقة أن تحليل التباين بثلاث اتجاهات (وn مشاهدة) يحتاج إلى حسابات معقدة وإلى عدد كبير من المشاهدات، ولكن يمكننا في بعض الأبحاث تنظيم وترتيب العوامل الداخلة في تحليل التباين الثلاثي وعرضها على شكل مربع يسمى المربع اللاتيني، وهو يستخدم كثيراً في الأبحاث العلمية كالأبحاث الزراعية والطبية والاقتصادية وغيرها، وهو يتألف من عدة عناصر هي :

- العامل الأول F_A ويأخذ g حالة منفصلة.
- العامل الثاني F_B ويأخذ أيضاً g حالة منفصلة.
- المربع اللاتيني الذي يتم تشكيله من تقاطع حالات العاملين F_A وF_B . وهو يتألف من g^2 خلية .

- جملة من طرائق المعالجة وعددها يساوي g طريقة أيضاً وذلك لتطبيقها على خلايا المربع اللاتيني، بحيث يتم تطبيق كل طريقة مرة واحدة في كل سطر، ومرة واحدة في كل عمود.
- متحول تابع X يقيس نتائج المعالجات السابقة في جميع الخلايا بواحدة قياس موحدة.

ولتمثيل أحد أشكال المربع اللاتيني ونفترض أن عدد حالات F_A و F_B يساوي $g = 3$ ، وأنه لدينا (3) طرائق هي $A\ B\ C$ لتطبيقها على تقاطعات تلك الحالات، مرة واحدة في كل سطر ومرة واحدة في كل عمود، وعندها يمكن أن يأخذ المربع اللاتيني الشكل التالي:

	F_{A1}	F_{A2}	F_{A3}
F_{B1}	A	B	C
F_{B2}	B	C	A
F_{B3}	C	A	B

الشكل (7− 2)

وهنا نلاحظ أن الشكل السابق (7− 2) للمربع اللاتيني هو أحد الأشكال الممكنة، وتم الحصول عليه بتطبيق تسلسل معين للطرائق المستخدمة هو $(A\ B\ C)$ على خلايا السطر الأول، ثم القيام بسحب عناصره إلى اليسار خطوة واحدة وتطبيقها على خلايا السطر الثاني (مع نقل A إلى الحجرة الأخيرة)، ثم سحب عناصره إلى اليسار خطوة واحدة وتطبيقها على خلايا السطر الثالث (مع نقل B إلى الحجرة الأخيرة)، وبذلك نحصل على ما يسمى الشكل القياسي للمربع اللاتيني، وهو المربع الذي يتم تشكيله بتدوير عناصر السطر الأول لتوزيعها على السطر الثاني ثم تدوير عناصر الثاني لتوزيعها على الثالث، وهكذا دواليك . وبذلك تكون عناصر الأسطر متناظرة مع عناصر الأعمدة. ولكن الشكل القياسي للمربع اللاتيني ليس وحيداً ,فهناك عدد من الأشكال التي تعطينا توزيعات مختلفة , حيث نجد أنه يمكننا ترتيب عناصر السطر الأول عشوائياً بـ !3 طريقة, ثم ترتيب العناصر المتبقية في العمود الأول عشوائياً (بعد السطر الأول) بـ !2 طريقة وتبقى الخلية الأخيرة والتي ترتب بطريقة واحدة فقط . وبذلك يكون عدد الأشكال الممكنة (حسب الأسطر أو الأعمدة) في الحالة التي يكون فيها $n = 3$ مساوياً لـ: $m = 3! * 2! * 1 = 12$, وهذه الأشكال الممكنة تعطينا المربعات اللاتينية التالية:

$$\begin{bmatrix} A & B & C \\ B & C & A \\ C & A & B \end{bmatrix} \begin{bmatrix} A & C & B \\ C & B & A \\ B & A & C \end{bmatrix} \begin{bmatrix} B & A & C \\ A & C & B \\ C & B & A \end{bmatrix} \begin{bmatrix} B & C & A \\ C & A & B \\ A & B & C \end{bmatrix} \begin{bmatrix} C & A & B \\ A & B & C \\ B & C & A \end{bmatrix} \begin{bmatrix} C & B & A \\ B & A & C \\ A & C & B \end{bmatrix}$$

$$\begin{bmatrix} A & B & C \\ C & A & B \\ B & C & A \end{bmatrix} \begin{bmatrix} A & C & B \\ B & A & C \\ C & B & A \end{bmatrix} \begin{bmatrix} B & A & C \\ C & B & A \\ A & C & B \end{bmatrix} \begin{bmatrix} B & C & A \\ A & B & C \\ C & A & B \end{bmatrix} \begin{bmatrix} C & A & B \\ B & C & A \\ A & B & C \end{bmatrix} \begin{bmatrix} C & B & A \\ A & C & B \\ B & A & C \end{bmatrix}$$

ولكننا في الأبحاث العلمية لا نحتاج إلى جميع هذه الأشكال، بل نقوم باختيار أحدها عشوائياً، وبذلك تكون نتيجة التجارب (طرائق المعالجة) في كل خلية هي عبارة عن متحول عشوائي معرف عليها.

وهكذا نجد أن المربع اللاتيني هو تحليل ثلاثي الاتجاه (بمشاهدة واحدة) وإن العوامل المعتمدة فيه هي:

- عامل الأسطر ونرمز له بـ F_A وعدد حالاته g حالة.
- عامل الأعمدة ونرمز له بـ F_B وعدد حالاته g حالة أيضاً.
- طرائق المعالجة وعددها g حالة أيضاً. وهي تطبق على خلايا المربع اللاتيني، بحيث يتم تطبيق كل طريقة مرة واحدة في كل سطر ومرة واحدة في كل عمود, وحسب أحد الأشكال الممكنة للمربع.

ونتيجة لتطبيق هذه الطرائق على الخلايا نحصل على g^2 قياساً للمتحول التابع X، وسنرمز لقيمة المتحول التابع X في الخلية $(k\,,\ell)$ بالرمز $x_{k\ell}$ ونضعها إلى جانب رمز الطريقة المطبقة في تلك الخلية، ثم نحسب مجاميع ومتوسطات الأسطر والأعمدة ونضعها في الهوامش.

وإذا أخذنا الحالة التي يكون لدينا فيها: $(g = 4)$ طرائق ويكون لكل من العاملين F_A و F_B أربعة حالات، فإننا سنحصل على الجدول التالي:

جدول (7−11): المربع اللاتيني 4×4

الأعمدة F_A / الأسطر F_B	1	2	3	4	المجموع X_k	المتوسط $\bar{X}_k$
1	A / x_{11}	B / x_{12}	C / x_{13}	D / x_{14}	X_1	$\bar{X}_1$
2	B / x_{21}	C / x_{22}	D / x_{23}	A / x_{24}	X_2	$\bar{X}_2$
3	C / x_{31}	D / x_{32}	A / x_{33}	B / x_{34}	X_3	$\bar{X}_3$
4	D / x_{41}	A / x_{42}	B / x_{43}	C / x_{44}	X_4	$\bar{X}_4$
المجموع X_ℓ	X'_1	X'_2	X'_3	X'_4	X	$\bar{X}$
المتوسط $\bar{X}_\ell$	$\bar{X}'_1$	$\bar{X}'_2$	$\bar{X}'_3$	$\bar{X}'_4$	$\bar{X}$	$\bar{X}$

حيث أن هذه المجاميع والمتوسطات تحسب من العلاقات التالية:

المتوسط في السطر k: $\bar{X}_k = \frac{1}{n}\sum_{\ell=1}^{g} x_{k\ell} = \frac{X_k}{g}$ ، اجمالي السطر k: $X_k = \sum_{\ell=1}^{g} x_{k\ell}$

(52 − 7) المتوسط في العمود ℓ: $\bar{X}_\ell = \frac{1}{n}\sum_{k=1}^{g} x_{k\ell} = \frac{X_\ell}{g}$ ، اجمالي العمود ℓ: $X_\ell = \sum_{k=1}^{g} x_{k\ell}$

المتوسط في الجدول ككل: $\bar{X} = \frac{X}{g*g} = \frac{\sum^g\sum^g x_{k\ell}}{g*g}$, الاجمالي الكلي $X = \sum_{k=1}^{g} X_k = \sum_{k=1}^{g} X_\ell = \sum_{k=1}^{g}\sum_{\ell=1}^{g} x_{k\ell}$

وهناك متوسطات من نوع آخر هي متوسطات قياسات X حسب طرائق المعالجة $(A\ B\ C\ D)$ ويتم حساب هذه المتوسطات لكل طريقة على حدة، وذلك بتتبع قيم X حسب كل طريقة ضمن المربع اللاتيني, ونرمز لهذه المتوسطات بالرموز $\bar{X}_t$، حيث t هو دليل الطريقة t، فمثلاً نجد أن متوسط قيم X

المقابلة للطريقة A يحسب من الخلايا التي تطبق فيها الطريقة A وهو يساوي (حسب الجدول السابق (7−11)) ما يلي:

$$\bar{X}_A = \frac{1}{4}[x_{11} + x_{24} + x_{33} + x_{42}] = \frac{X_A}{g}$$

وكذلك نجد أن متوسطات الطرائق الأخرى تساوي:

$$\bar{X}_B = \frac{1}{4}[x_{12} + x_{21} + x_{34} + x_{43}] = \frac{X_B}{g}$$

$$\bar{X}_C = \frac{1}{4}[x_{13} + x_{22} + x_{31} + x_{44}] = \frac{X_C}{g} \qquad (53-7)$$

$$\bar{X}_D = \frac{1}{4}[x_{14} + x_{23} + x_{32} + x_{41}] = \frac{X_D}{g}$$

ويمكن تسهيل حساب هذه المتوسطات بتنظيم جدول خاص بذلك، يتضمن تبويب قياسات X حسب طرائق المعالجة، وحسب الأسطر (أو حسب الأعمدة) كما يلي:

جدول (7−12): حساب المتوسطات حسب طرائق المعالجة

الطريقة t / الأسطر	A	B	C	D	X_i
1	x_{11}	x_{12}	x_{13}	x_{14}	X_1
2	x_{24}	x_{21}	x_{22}	x_{23}	X_2
3	x_{33}	x_{34}	x_{31}	x_{32}	X_3
4	x_{42}	x_{43}	x_{33}	x_{41}	X_4
المجموع	X_A	X_B	X_C	X_D	X
المتوسط $\bar{x}_t$	$\bar{X}_A$	$\bar{X}_B$	$\bar{X}_C$	$\bar{X}_D$	

وهنا يشترط على هذه القياسات في كل عمود (أو سطر) ان تكون مستقلة عن بعضها البعض ، وخاضعة للتوزيع الطبيعي ولها تباين موحد مساوٍ لـ σ^2 .

ولنفترض الآن أن التوقع الرياضي لقيم X في الخلية (k , ℓ) من المربع يساوي $\mu_{k\ell}$، وأن التوقع العام لقيم X في كامل المربع بالرمز μ . وبما أن القياسات مستقلة فإن التفاعلات الثنائية للعوامل تكون معدومة، كما أن التفاعل الثلاثي للعوامل الثلاثة معاً يكون معدوماً أيضاً.

وبناء على ذلك يمكننا أن نصيغ نموذج تحليل المربع اللاتيني لهذه العوامل الثلاثة المستقلة كما يلي:

$$\mu_{k\ell} = \mu + \propto_k + \beta_\ell + \gamma_t \qquad (54-7)$$

حيث أن: $\propto_k$ هو تأثير السطر k و β_ℓ تأثير العمود ℓ، و γ_t تأثير الطريقة t، وبما أن المتوسط العام في العينات هو تقدير لـ μ في المجتمع، فإنه يمكننا التعبير عن قيمة X في كل خلية كما يلي:

$$x_{k\ell} = \bar{X} + \propto_k + \beta_\ell + \gamma_t + e_{k\ell} \qquad (54\,a-7)$$

حيث أن: $e_{k\ell}$ هو حد الخطأ العشوائي ويخضع للتوزيع الطبيعي $N(0\,,\sigma^2)$، وهكذا نجد أن مجموع مربعات الانحرافات عن μ يساوي:

$$\sum_{k=1}^{g}\sum_{\ell=1}^{g}(x_{k\ell}-\mu)^2 \qquad (55-7)$$

ولإظهار تأثير العوامل الثلاثة نستبدل μ بتقديره $\bar{X}$ ثم نضيف ونطرح المتوسطات $\bar{X}_k$ و $\bar{X}_\ell$ و $\bar{X}_\ell$ ونكتب المجموع السابق كما يلي ($\bar{X}_t$ متوسط الطريقة t):

$$\sum_{k=1}^{g}\sum_{\ell=1}^{g}(x_{k\ell}-\bar{X})^2=\sum_{k=1}^{g}\sum_{\ell=1}^{g}[(x_{k\ell}-\bar{X}_k-\bar{X}_\ell-\bar{X}_t+2\bar{X})+(\bar{X}_k-\bar{X})+(\bar{X}_\ell-\bar{X})+(\bar{X}_t-\bar{X})]^2 \quad \mathbf{(56-7)}$$

ثم نقوم بتربيع الحدود التي داخل القوس المتوسط، ولكن نظراً لاستقلال العوامل المستخدمة في التحليل فإن مجاميع الجداءات الثنائية تكون معدومة (يمكن البرهان على ذلك لكل حد على حدة . انظر المرجع Dugue P.282) .

ونتيجة بعض الإصلاحات نحصل على أن مجموع مربعات الانحرافات الإجمالي يساوي:

$$\sum_{k=1}^{g}\sum_{\ell=1}^{g}(x_{k\ell}-\bar{X})^2=g\sum_{k=1}^{g}(\bar{X}_k-\bar{X})^2+g\sum_{\ell=1}^{g}(\bar{X}_\ell-\bar{X})^2+g\sum_{t=1}^{g}(\bar{X}_t-\bar{X})^2+$$
$$+\sum_{k=1}^{g}\sum_{\ell=1}^{g}(x_{k\ell}-\bar{X}_k-\bar{X}_\ell-\bar{X}_t+2\bar{X})^2 \qquad \mathbf{(57-7)}$$

وباستخدام نفس الرموز السابقة نكتب المجاميع السابقة كما يلي:

$$SST=SSA+SSB+SSt+SSE \qquad (58-7)$$

وإن درجات الحرية التي تقابلها تساوي ما يلي:

$$g^2-1=(g-1)+(g-1)+(g-1)+(g-1)\,(g-2) \qquad (59-7)$$

وبعد حساب هذه المجاميع نضعها في جدول تحليل التباين الثلاثي كالتالي:

جدول (7–13): تحليل التباين في المربع اللاتيني:

مصدر التباين	مجموع مربعات الانحرافات	درجة الحرية	متوسط المربعات	F المحسوبة
F_A	$SSA=$	$g-1$	$MSA=\dfrac{SSA}{g-1}$	$F_A=\dfrac{MSA}{MSE}$
F_B	$SSB=$	$g-1$	$MSB=\dfrac{SSB}{g-1}$	$F_B=\dfrac{MSB}{MSE}$
طرائق المعالجة t_r	$SSt_r=$	$g-1$	$MSt=\dfrac{SSt_r}{g-1}$	$F_t=\dfrac{MSt}{MSE}$
الخطأ العشوائي	$SSE=$	$(g-1)\,(g-2)$	$MSE=\dfrac{SSE}{(g-1)\,(g-2)}$	
الاجمالي	SST	g^2-1	——	

أما بالنسبة للفرضيات فهي تنطلق من النموذج (7–54) ونكتبها كما يلي:

$$H_0: \begin{cases} \propto_k = 0 & : K : 1\ 2\ 3 \ldots g \\ \beta_\ell = 0 & : \ell : 1\ 2\ 3 \ldots g \\ \gamma_t = 0 & : t : 1\ 2\ 3 \ldots g \end{cases} \qquad (60-7)$$

أما الفرضية البديلة H_1 فتكتب كما يلي:

$$H_1: \begin{cases} \propto_k \neq 0 & \text{من أجل } k \text{ واحدة على الأقل} \\ \beta_\ell \neq 0 & \text{من أجل } \ell \text{ واحدة على الأقل} \\ \gamma_t \neq 0 & \text{من أجل } t \text{ واحدة على الأقل} \end{cases} \qquad (61-7)$$

وهنا نفترض أن تصميم المربع اللاتيني يعتبر أن تأثيرات الأسطر وتأثيرات الأعمدة ثابتة، رغم أنها تستخدم للتحكم في مصادر الاختلاف . لذلك فإن تأثيراتها يجب أن تحقق العلاقتين التاليتين:

$$\sum_{k=1}^{g} \propto_k = 0 \qquad \sum_{\ell=1}^{g} \beta_\ell = 0 \qquad (62-7)$$

أما تأثيرات طرائق المعالجة γ_t فإما أن تكون ثابتة أو عشوائية.

فإذا كانت γ_t ثابتة فإن تأثيراتها تقدر على أنها انحرافات عن المتوسط العام $\bar{X}$ ، ولذلك فإنها يجب أن تحقق العلاقة التالية:

$$\sum_{t=1}^{g} \gamma_t = 0 \qquad \left(\text{إذا كانت } \gamma_t \text{ ثابتة}\right) \qquad (63-7)$$

أما إذا كانت γ_t عشوائية فإننا نفترض ونتحقق من أنها خاضعة للتوزيع الطبيعي $N(0, \sigma^2)$، ثم نتعامل معها بأسلوب آخر (اختبار Tukey) . وهنا نشير إلى أن الفرضية الأساسية التي نريد اختبارها هي فرضية العدم التالية :

$$H_0: \gamma_1 = \gamma_2 = \cdots \gamma_g = 0 \qquad (64-7)$$

مقابل الفرضية البديلة:

$$H_1: \gamma_t \neq 0 \qquad \text{من أجل } t \text{ واحدة على الأقل}$$

ولاختبار الفرضية $H_0: \gamma_1 = \gamma_2 = \cdots \gamma_g = 0$ نقوم بحساب مجاميع مربعات الانحرافات ومتوسطاتها ونضعها في جدول تحليل التباين، ونبدأ بحساب مؤشر الاختبار لتأثير المعالجات من العلاقة:

$$F_t = \frac{\frac{SSt}{g-1}}{\frac{SSE}{(g-1)(g-2)}} = \frac{MSt_r}{MSE} \qquad (65-7)$$

ثم نقارن هذه القيمة المحسوبة F_t مع القيمة الحرجة لـ F والمقابلة لمستوى الدلالة $\propto = 0.05$ ولدرجتي الحرية $v_1 = (g-1)$ ، $v_2 = (g-1)(g-2)$ ونتخذ القرار كما يلي:

إذا كانت $F_t \leq F_{v_1\, v_2}(\propto)$ نقبل الفرضية H_0، والتي تقول أن تأثيرات طرائق المعالجة معدومة أو إنها غير معنوية والعكس بالعكس.

أما بالنسبة لاختبار تأثير العاملين F_A و F_{B} نقوم بحساب قيمتي المؤشرين:

$$F_A = \frac{MSA}{MSE}$$
$$F_B = \frac{MSB}{MSE} \qquad (66-7)$$

ثم نقوم بمقارنة كل منهما مع القيمة الحرجة $F_{v_1\, v_2}(\propto)$، ونقبل الفرضية H_0 حول $\propto_k = 0$ إذا كان $F_A \leq F(\propto)$، كما نقبل الفرضية H_0 حول $\beta_\ell = 0$ إذا كان $F_B \leq F(\propto)$ والعكس بالعكس.

ولكن المؤشرين F_A و F_B يعتبران مؤشرين تقريبيين، لذلك نحسب الكفاءة النسبية للأسطر وللأعمدة من العلاقتين التاليتين:

$$RE\ \left(\text{للأسطر}\right) = \frac{MSA + (g-1)MSE}{g * MSE}100 \qquad (67-7)$$

$$RE\ \left(\text{للأعمدة}\right) = \frac{MSB + (g-1)MSE}{g * MSE}100 \qquad (68-7)$$

ولتقدير الكفاءة الكلية للمربع اللاتيني نستخدم العلاقة:

$$RE\ \left(\text{للمربع}\right) = \frac{MSA + MSB + (g-1)MSE}{(g+1)MSE} \qquad (69-7)$$

ملاحظة إذا كانت قيمة إحدى المشاهدات $x_{k\ell}$ مفقودة، فإنه لا يمكننا تطبيق أسلوب المربع للاتيني بدونها، وحتى نتابع العمل علينا أن نقوم بتقديرها من العلاقة:

$$\tilde{x}_{k\ell} = \frac{g(X_k + X_\ell + X_t) - 2X}{(g-1)(g-2)} \qquad (70-7)$$

وهنا علينا أن نخفض درجات الحرية للأخطاء بمقدار درجة واحدة وتصبح كما يلي:

$$(g-1)(g-2) - 1$$

وأن نخفض درجات الحرية لإجمالي المربعات درجة واحدة أيضاً فتصبح : $(g^2 - 2)$

ولتقدير الخطأ المعياري لتقدير المتوسط $\bar{x}_t$ نستخدم العلاقة :

$$S_{\bar{x}_t} = \sqrt{\frac{MSE}{g}} \qquad (71-7)$$

ثم ننشأ مجال الثقة له ذي الاحتمال $(1 - \propto)$ من العلاقة:

$$\bar{x}_t \pm t_{(g-1)(g-2)}\left(\frac{\propto}{2}\right) * S_{\bar{x}_k} \qquad (72-7)$$

كما يمكننا حساب الخطأ المعياري للفرق بين متوسطي طريقتين $(\bar{x}_{t1} - \bar{x}_{t2})$ من العلاقة:

$$S_{(\bar{x}_{t1} - \bar{x}_{t2})} = \sqrt{\frac{MSE}{g} + \frac{MSE}{g}} = \sqrt{\frac{2MSE}{g}} \qquad (73-7)$$

ثم إنشاء مجال الثقة له ذي الاحتمال $(1-\propto)$ من العلاقة:

$$(\bar{x}_{t1}-\bar{x}_{t2}) \pm t_{(g-1)(g-2)}\left(\frac{\propto}{2}\right) * S_{(\bar{x}_{t1}-\bar{x}_{t2})} \qquad (74-7)$$

مثال (7–5): لدراسة أداء (6) عمال من عمال شركة النسيج، نرمز لهم بالرموز (A, B, C, D, E, F) تم اعتماد المتحول التابع X الذي يعبر عن إنتاجية العامل في اليوم من القماش (م^2 / يوم) . ولإجراء التجارب تم تجهيز (6) آلات نسيج مختلفة، على أن يتم توزيع هؤلاء العمال على تلك الآلات عشوائياً في كل يوم ولمدة (6) أيام (أيام العمل في الأسبوع)، وبحيث يعمل كل عامل مرة واحدة على آلة مختلفة في كل يوم من ايام العمل في الأسبوع (مرة واحدة في كل عمود ومرة واحدة في كل سطر).

والمطلوب دراسة تأثير كل من الأيام D والآلات M والعمال L على إنتاجية هؤلاء العمال، إذا علمت أن توزيعات هؤلاء العمال وإنتاجياتهم اليومية كانت بعد تنفيذ الدراسة كما يلي:

جدول (7–14): بيانات المربع اللاتيني للمثال (فرضية)

الآلات M_ℓ / الأيام D_k	M_1	M_2	M_3	M_4	M_5	M_6	المجموع X_k	المتوسط $\bar{X}_k$
السبت D_1	B 28.7	E 28.4	D 25.4	C 30.7	A 30.6	F 30.9	$X_1 = 174.7$	29.12
الأحد D_2	F 31.4	C 30.1	B 27.4	E 26.8	D 29.8	A 29.8	$X_2 = 175.3$	29.22
الاثنين D_3	D 29.4	F 29.7	A 30.4	B 22.0	E 24.1	C 32.9	$X_4 = 163.0$	28.08
الثلاثاء D_4	A 29.6	B 21.8	E 22.5	F 30.0	C 39.6	D 28.5	$X_3 = 168.5$	27.17
الأربعاء D_5	C 25.8	D 21.9	F 23.1	A 24.3	B 20.7	E 17.7	$X_5 = 133.5$	22.25
الخميس D_6	E 18.1	A 23.6	C 22.5	D 20.2	F 23.7	B 18.9	$X_6 = 127.0$	21.17
المجموع X_ℓ	163.0	155.3	151.3	154.0	159.5	158.7	$X = 942.0$	—
المتوسط $\bar{X}_\ell$	27.17	25.88	25.22	25.67	26.58	26.45	—	$\bar{X} = 26.17$

ولحساب إنتاجية كل عامل على حدة علينا أن نتتبع مقدار إنتاجيته على كل آلة وفي كل يوم . لذلك نتتبع إنتاجية هؤلاء العمال حسب الآلات ونصمم جدولاً خاصاً لتبويبها من جديد حسب العمال والآلات فنحصل من الجدول (7–14) السابق على الجدول التالي:

جدول (7–15) إنتاجية العمال حسب الآلات

العمال / الآلات	A	B	C	D	E	F	المجموع X_ℓ	المتوسط $\bar{X}_\ell$
M_1	29.6	28.7	25.8	29.4	18.1	31.4	163.0	27.17
M_2	23.6	21.8	30.1	21.9	28.4	29.7	155.5	25.88
M_3	30.4	27.4	22.5	25.4	22.5	23.1	151.3	25.22
M_4	24.3	22.0	30.7	20.2	26.8	30.0	154.0	25.67
M_5	30.6	20.7	30.6	29.8	24.1	23.7	159.5	26.58
M_6	29.8	18.9	32.9	28.5	19.7	30.9	158.7	26.45
X_t	168.3	139.5	172.6	155.2	137.7	168.8	942.0	—
$\bar{X}_t$	28.05	23.25	28.77	25.87	22.95	28.13	—	26.17

ومن السطر الأخير للجدول (7–15) نلاحظ أن إنتاجية هؤلاء العمال تختلف من عامل لآخر، وإن أحسنها هي إنتاجية العمال C ثم F ثم A، وإن أسوأها هي إنتاجية العمال E ثم B ثم D . وكذلك يمكننا أن نبوب إنتاجية هؤلاء العمال حسب الأيام فنحصل من الجدول (7–14) على جدول مشابه للجدول (7–15) السابق.

إلا أن حساب الجدول (7–15) يغني عنه في الحسابات اللاحقة . والآن نعتمد النموذج النظري التالي:

$$x_{k\ell} = \bar{X} \pm \propto_k + B_\ell + \gamma_t + e_{k\ell}$$

ثم نضع الفرضيات كما يلي:

$$H_0: \begin{cases} \propto_1 = \propto_2 = \cdots = \propto_6 = 0 \\ B_1 = B_2 = \cdots = B_6 = 0 \\ \gamma_1 = \gamma_2 = \cdots = \gamma_6 = 0 \end{cases}$$

$$H_1: \begin{cases} \propto_k \neq 0 & \text{من أجل } k \text{ واحدة على الأقل} \\ B_\ell \neq 0 & \text{من أجل } \ell \text{ واحدة على الأقل} \\ \gamma_t \neq 0 & \text{من أجل } t \text{ واحدة على الأقل} \end{cases}$$

ولاختبار هذه الفرضيات نقوم بحساب مجاميع مربعات الانحرافات SST وSSA وSSB وSSt وSSE من بيانات الجدولين (7–14) و(7–15) ونستخدم العلاقات التالية:

$$SST = \sum_{k=1}^{6}\sum_{\ell=1}^{6}(x_{k\ell} - \bar{X})^2 = \sum_{k=1}^{6}\sum_{\ell=1}^{6} x_{k\ell}^2 - \frac{X^2}{g^2}$$

$$SST = [(28.7)^2 + (28.4)^2 + \cdots (23.7)^2 + (18.9)^2] - \frac{(942)^2}{36}$$

$$SST = 25299.10 - 24649 = 650.10$$

ثم نقوم بحساب مجموع المربعات للأيام D فنجد من مجاميع الأسطر في الجدول (7–14) أن:

$$SSA = \sum_{k=1}^{6} X_k^2 - \frac{X^2}{g^2} = [(174.7)^2 + (175.3)^2 + \cdots (127.0)^2] - \frac{(942)^2}{36} = 378.11$$

ثم نقوم بحساب مجموع المربعات للآلات فنجد من مجاميع الأعمدة في الجدول (7−14) أن:

$$SSB=\sum_{\ell=1}^{6}X_{\ell}^{2}-\frac{X^{2}}{g^{2}}=[(163.3)^{2}+(155.5)^{2}+\cdots(158.7)^{2}]-\frac{(942)^{2}}{36}=14.81$$

ثم نقوم بحساب مجموع المربعات للعمال اعتماداً على أعمدة الجدول (7-15) فنجد أن:

$$SSt=\sum_{t=1}^{6}X_{t}^{2}-\frac{X^{2}}{g^{2}}=[(168.3)^{2}+(139.5)^{2}+\cdots(168.8)^{2}]-\frac{(942)^{2}}{36}=199.36$$

ثم نقوم بحساب SSE من العلاقة:

$$SSE=SST-SSA-SSB-SSt$$

$$SSE=650.10-378.4-14.81-199.36=57.82$$

وأخيراً نضع نتائج هذه الحسابات مع درجات حرياتها في جدول تحليل التباين الثلاثي فنحصل على أن:

جدول (7−16): تحليل التباين للمربع اللاتيني:

مصدر التباين	مجموع المربعات	درجة حرية	متوسط المربعات	قيمة F المحسوبة
الأسطر (الأيام)	$SSA=378.11$	$g-1=5$	75.62	$F_A=26.16$
الأعمدة (الآلات)	$SSB=14.8$	$g-1=5$	2.96	$F_B=1.02$
المعالجات (العمال)	$SSt=199.36$	$g-1=5$	39.87	$F_t=13.79$
الأخطاء	$SSE=57.82$	$(g-1)(g-2)=20$	2.89	ـــــــ
الإجمالي	$SST=650.10$	$g^2-1=35$	ـــــــ	ـــــــ

ثم نقوم بإيجاد قيمة F الحرجة الموافقة لدرجتي حرية $v_1=5$, $v_2=20$ ولمستوى دلالة $\propto=$ 0.05 فنجد أن : $F_{v_1,v_2}(\propto)=F_{5,20}(0.05)=2.71$

وهي نفسها لجميع العوامل، لذلك نقارن F_A و F_B و F_t معها، فنجد ما يلي:

1− بالنسبة للعمال نجد أن: $F_t>F(\propto)$، لذلك نرفض فرضية العدم $H_0:\gamma_1=\gamma_2=\cdots=\gamma_6=0$ ونقبل الفرضية البديلة H_1 ، التي تقول أن الإنتاجية X تتأثر بالعمال (وهذا أمر واضح من السطر الأخير في الجدول (7−15).

2− بالنسبة للأيام (الأسطر) نجد أن $F_A>F(\propto)$، لذلك نرفض فرضية العدم $H_0:\propto_1=\propto_2=\cdots=\propto_6=0$ ونقبل الفرضية البديلة H_1، التي تقول أن الإنتاجية X تتأثر بالأيام (وهذا أمر واضح من العمود الأخير في الجدول (7−10).

3− بالنسبة للآلات (الأعمدة) نجد أن $F_B<F(\propto)$ لذلك نقبل فرضية العدم $H_0:B_1=B_2=\cdots=B_6=0$ ونستنتج أن الآلات لا تؤثر على إنتاجية العمال.

ولتقدير كفاءة هذا التصميم نقوم بحساب الكفاءة النسبية للمربع ككل فنجد أن:

$$RE(Squere)=\frac{MSA+MSB+(g-1)MSE}{(g+1)MSE}100=\frac{75.62+2.96+5*2.89}{7*2.89}100=459.86\%$$

وهذا يدل على كفاءة عالية بالنسبة للقطاعات العشوائية التامة.

كما يمكننا حساب الكفاءة النسبية للأسطر (بدون الأعمدة) من العلاقة:

$$RE(row) = \frac{MSA + (g-1)MSE}{g * MSE}100 = \frac{75.62 + 5 * 2.89}{6 * 2.89}100 = 519.43\%$$

وهذا يدل على كفاءة عالية بالنسبة للقطاعات العشوائية التامة.

وكذلك نقوم بحساب الكفاءة النسبية للأعمدة (بدون الأسطر) من العلاقة:

$$RE(column) = \frac{MSB + (g-1)MSE}{g * MSE}100 = \frac{2.96 + 5 * 2.89}{6 * 2.89}100 = 100.40\%$$

وهذا يدل على أن كفاءة الأعمدة (الآلات) بقيت تساوي نفسها 100% ولم تؤثر على إنتاجية العمال.

7-5: تحليل التباين المشترك (تحليل التغايرANCOVA):

7-5-1: تمهيد:

لقد استعرضنا في النماذج السابقة لتحليل التباين النماذج التي تدرس تغيرات متحول تابع X (Dependent variable)، الناتجة عن متغير أو متغيرات وصفية مستقلة (Independent)، مثل المجتمعات أو المعالجات أو القطاعات.

ولكن إذا كان التابع X مرتبطاً بمتحول كمي آخر Y غير مرغوب به، ولكنه ملازم لـ X ولا نستطيع التحكم فيه، ولكننا نريد التخلص من تأثيره على التابع X . فإننا نلجأ إلى إجراء تحليل التباين المشترك (التغاير ANCOVA) لفصل تأثيرات Y غير المرغوب فيها عن تغيرات X ونتبع في ذلك إحدى الطريقتين التاليتين:

1- إزالة اختلافات X المرتبطة بـ Y من قياسات X (إذا كان ذلك ممكناً) وعندها نحصل على قياسات صافية لـ X . ثم نقوم بإجراء تحليل التباين على القياسات الصافية فنحصل على اختبارات قوية (ولكن هذه الطريقة قد لا تكون ممكنة في معظم الحالات).

2- تعديل متوسطات المعالجات في المجتمعات، بحيث يتم طرح قيم موحدة للمتغير المستقل Y منها، وبالتالي نحصل على طريقة عادلة لمقارنة قيم متوسطات X في تلك المجتمعات المختلفة.

فمثلاُ لدراسة تغيرات الانفاق الشهري X لطلاب الجامعة حسب الجنس (مجتمع الذكور ومجتمع الإناث). نلاحظ أن ذلك الانفاق لا يتأثر بنوع الجنس فقط، بل يتأثر أيضاً بمقدار الدخل الشهري المخصص للطالب والذي سنرمز له بـ Y . لذلك علينا أن نقوم بعزل تأثير الدخل Y على X ، وذلك حتى نتمكن من إجراء مقارنة عادلة لمتوسطي الإنفاق X حسب الجنس دون تأثير الدخل Y عليهما.

وللتخلص من تأثيرات المتحول الإضافي Y على قيم X نلجأ إلى تحليل التباين المشترك (التغاير ANCOVA)، وإذا اجرينا تحليل التباين البسيط (ANOVA) بدون عزل تأثير Y، فإن ذلك سيؤدي إلى تضخيم الخطأ التجريبي، ويصبح من الصعب اكتشاف الفروقات الحقيقية بين المجتمعات. وبذلك نجد أن المهمة الرئيسية لتحليل التغاير هي تصغير قيمة الخطأ التجريبي.

وتتضمن المراجع المختصة عدة أنواع لتحليل التغاير هي:

- تحليل التغاير في تصميم العشوائية التامة (CRD).
- تحليل التغاير في تصميم القطاعات العشوائية (RCBD).
- تحليل التغاير في تصميم التجارب العاملية 2^k $(2^k FD)$.

وسنقتصر في منشورنا هذا على النوع الأول لأنه الأكثر انتشاراً والأسهل تطبيقاً.

7−5−2: تحليل التغاير في تصميم العشوائية التامة (CRD) باتجاه واحد (one-way ANCOVA) .

يمكن النظر إلى تحليل التغاير على أنه تعديل لتحليل التباين البسيط باتجاه واحد (ANOVA)، وذلك بعد إضافة المتحول Y إلى النموذج الرياضي ثم إعادة صياغة النموذج بحيث يتم عزل تأثير Y منه .

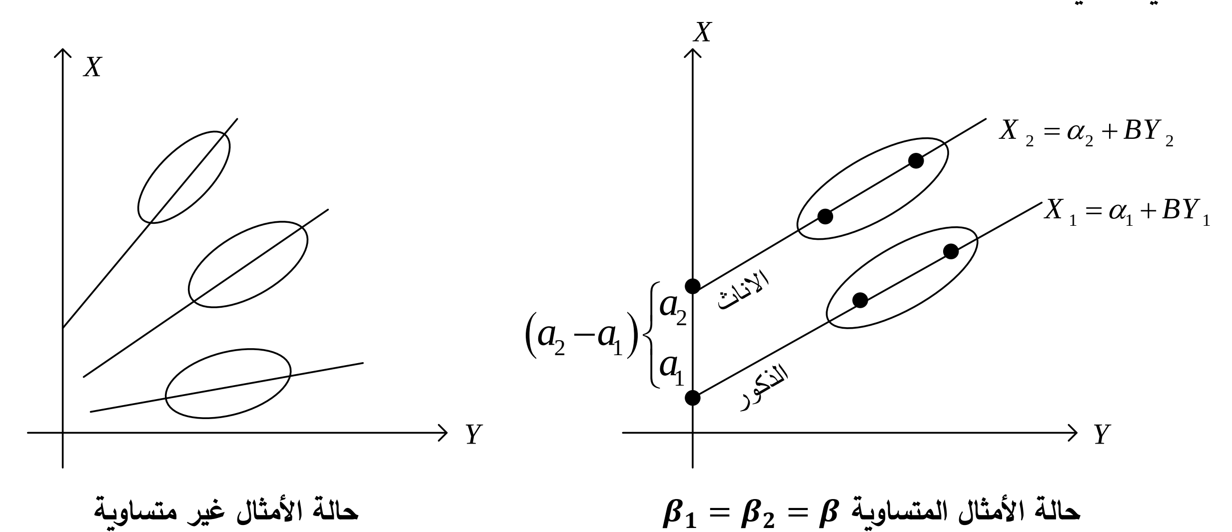
وبصورة عامة نفترض أنه لدينا g مجتمعاً، تؤثر على متحول تابع X ، وأن X يترافق مع متحول إضافي كمي Y ، وأن X يرتبط مع Y في كل مجتمع k بعلاقة انحدار خطية من الشكل:

$$X_k = a_k + \beta_k Y_k = a_k + \beta Y_k \qquad : k: 1\ 2\ 3 \dots g \qquad (75-7)$$

ولتسهيل صياغة النموذج الرياضي المشترك افترضنا أن قيم الأمثال β_k متساوية في جميع تلك المجتمعات، أي إننا نفترض أو نشترط أن يكون :

$$\beta_1 = \beta_2 \dots = \beta_g = \beta \qquad (75\,a-7)$$

وإن هذا يعني أن ميول مستقيمات هذه العلاقات متساوية في جميع المجتمعات المدروسة، وترسم الشكل اليميني التالي :



حالة الأمثال غير متساوية

حالة الأمثال المتساوية $\beta_1 = \beta_2 = \beta$

الشكل (7−3): حالات علاقة x بـ Y

كما نفترض أن حجوم العينات المسحوبة من تلك المجتمعات متساوية وتساوي $n_k = r$، ومن جهة ثانية نجد أنه يمكننا صياغة نموذج تحليل التباين البسيط (ANOVA) للمتحول التابع X حسب (7−7) كما يلي:

$$x_{ki} = \mu_x + \propto'_k + e'_{ki} \qquad (76-7)$$

حيث أن: $i = 1\ 2\ 3 \dots r$ و $k = 1\ 2\ 3 \dots g$

وحيث أن: μ_x هو توقع X، وإن: $\propto'_k$ هو مقدار تأثير المجتمع k على X، وهو يحقق الشرط التالي: $\sum_{k=1}^{g} \propto'_k = 0$.

وأن: e'_{ki} هي حدود الخطأ العشوائي (البواقي) ويشترط فيها أن تكون مستقلة وخاضعة للتوزيع الطبيعي $N(0\,,\sigma^2)$.

ومن جهة ثالثة نجد أنه يمكننا صياغة النموذج الرياضي لتحليل التباين (ANOVA) للمتحول الإضافي Y حسب (7–7) كما يلي:

$$y_{ki} = \mu_y + \propto''_k + e''_{ki} \qquad (77-7)$$

حيث أن: $i = 1\ 2\ 3 \dots r$ و $k = 1\ 2\ 3 \dots g$

وحيث أن: μ_y هو توقع Y، وإن: $\propto''_k$ هو مقدار تأثير المجتمع k على Y، وهو يحقق الشرط التالي: $\sum_{k=1}^{g} \propto''_k = 0$.

وأن: e''_{ki} هي حدود الخطأ العشوائي (البواقي) ويشترط فيها أن تكون مستقلة وخاضعة للتوزيع الطبيعي $N(0\,,\sigma^2)$.

وبعدها نضرب طرفي العلاقة (7–77) بقيمة الأمثال β (بعد حسابها من(7–75))، ونطرح أطرافها من أطراف العلاقة (7–76) فنحصل على ما يلي:

$$(x_{ki} - \beta y_{ki}) = (\mu_x - \beta\mu_y) + (\propto'_k - \beta \propto''_k) + (e'_{ki} - \beta e''_{ki}) \qquad (78-7)$$

وإذا رمزنا للطرف الأيسر بـ $Z_{ki} = (x_{ki} - \beta y_{ki})$ وأخذنا توقعه نجد أن:

$$E(Z_{ki}) = E(x_{ki}) - \beta E(y_{ki}) = \mu_x - \beta\mu_y = \mu_z \qquad (79-7)$$

وهكذا يظهر لدينا متحول جديد هو: $Z = X - \beta Y$

وأن توقعه: $\mu_z = \mu_x - \beta\mu_y$

وإذا رمزنا للحدود الأخرى في (7–78) بالرموز التالية:

$\propto_k = \propto'_k - \beta \propto''_k$ و $e_{ki} = e'_{ki} - \beta e''_{ki}$ فإننا نحصل على النموذج التالي:

$$Z_{ki} = \mu_z + \propto_k + e_{ki} \qquad (80-7)$$

حيث أن: $i = 1\ 2\ 3 \dots r$ و $k = 1\ 2\ 3 \dots g$ وهو نموذج تحليل التباين البسيط لـ Z .

حيث أن: μ_z هو توقع المتحول Z، وأن: $\propto_k$ هو تأثير المجتمع k على Z وهو يحقق الشرط التالي: $\sum_{k=1}^{g} \propto_k = 0$.

وأن: e_{ki} هي حدود الخطأ العشوائي الجديدة، وهي حدود مستقلة وتخضع للتوزيع الطبيعي $N(0\,,\sigma^2)$.

وهكذا نكون قد توصلنا إلى نفس الصيغة الرياضية التي يأخذها نموذج ANOVA للمتحول الجديد Z.

ولإظهار المتحول Y في النموذج (7–80) نعيد الرموز إلى أصولها وندمج الحدود المتشابهة، فنجد أن العلاقة (7–80) تأخذ الشكل التالي:

$$(x_{ki} - \beta y_{ki}) = (\mu_x - \beta\mu_y) + \propto_k + e_{ki}$$

وبعد الإصلاح نحصل على الصيغة التي تعطينا أية قيمة x_{ki} كما يلي:

$$x_{ki} = \mu_x + \ltimes_k + \beta(y_{ki} - \mu_y) + e_{ki} \qquad (81-7)$$

حيث أن: $k = 1\ 2\ 3 \dots g$ و $i = 1\ 2\ 3 \dots r$.

وهي صيغة نموذج تحليل التغاير ANCOVA للمتحول X المرتبط بالمتحول الإضافي Y. وهي الصيغة الأكثر انتشاراً .

ويمكننا الحصول على صيغة أخرى لـ (7−81) إذا قمنا بكتابتها كما يلي:

$$x_{ki} = (\mu_x + \beta\mu_y) + \ltimes_k + \beta y_{ki} + e_{ki}$$

وإذا رمزنا للمقدار $\mu' = \mu_x + \beta\mu_y$ نحصل على الصيغة التالية:

$$x_{ki} = \mu' + \ltimes_k + \beta y_{ki} + e_{ki} \qquad (82-7)$$

حيث أن: $k = 1\ 2\ 3 \dots g$ و $i = 1\ 2\ 3 \dots r$ ، وحيث أن: μ' هو مقدار ثابت يعبر عن توقع متحول ثالث وليس عن توقع X، لأن توقع X يساوي $\mu_x = \mu' + \beta\mu_y$.

وهكذا نكون قد توصلنا إلى نموذج تحليل التغاير للمتحول X المرتبط بالمتحول الإضافي Y، وهو يأخذ إحدى الصيغتين المتكافئتين (7−81) و(7−82).

وأخيراً نشير إلى أنه لتطبيق هذا النموذج يشترط على عناصره أن تحقق افتراضات تحليل التباين ANOVA وافتراضات الانحدار البسيط ونلخصها بما يلي:

- الافتراضات الموضوعة على نموذج ANCOVA :

1− أن تكون العينات المسحوبة عشوائية ومستقلة، وأن تكون حجومها متساوية وتساوي: $n_1 = n_2 = \cdots = n_g = r$.

2− أن يكون X وY متحولان كميان وخاضعان في كل مجتمع k للتوزيع الطبيعي بتوقعين μ_{kx} و μ_{ky} على الترتيب .

3− أن تكون تباينات التابع X في جميع المجتمعات متساوية وتساوي σ^2 .

4− أن تكون قياسات المتحول الإضافي Y نهائية، ولا تتضمن أخطاءً في القياس ولا تتأثر بالمجتمعات أو المعالجات .

5− أن تكون العلاقات بين X وY في جميع المجتمعات خطية وتأخذ الشكل التالي: $X_k = a_k + \beta_k Y_k$ وتكون $\beta_k \neq 0$.

6− أن تكون قيم جميع الأمثال β_k متساوية في جميع المجتمعات المدروسة أي $\beta_k = \beta$، أي أن تكون المستقيمات متوازية كما في الشكل (7−3) السابق .

7− أن يكون مجموع تأثيرات المجتمعات على X معدوماً . أي أن يكون: $\sum_{k=1}^{g} \ltimes_k = 0$.

8− أن تكون الأخطاء e_{ki} أو البواقي الناتجة عن النموذج مستقلة وخاضعة للتوزيع الطبيعي بتوقع معدوم وتباين ثابت σ^2 ، أي خاضعة لـ $N(0\,,\sigma^2)$.

ونلاحظ من هذه الافتراضات أن الافتراض الرابع ينص على أن المتغير الإضافي Y لا يتأثر بالمجتمعات (المعالجات) وهو أهم افتراض في تحليل التغاير ، لأنه إذا كان Y يتأثر بالمجتمعات، فإن معاملات الانحدار β_k ستختلف من مجتمع لآخر، وهذا يؤدي إلى عدم ثبات قيم تلك المعاملات في المجتمعات المختلفة، وهذا يخل في الافتراض السادس $(\beta_k = \beta)$ ويجعل النموذج (6−81) غير صالح للتطبيق على تلك المجتمعات.

ونستنتج مما سبق عند تطبيق ANCOVA أنه يجب علينا قبل كل شيء التحقق من أن Y لا يتأثر بالمجتمعات، وذلك بإجراء تحليل التباين ANOVA على Y بمفرده، فإذا كانت النتيجة قبول فرضية العدم H_0 (لا توجد فروقات لـ Y بين المجتمعات)، فإننا نتابع العمل للتحقق من الافتراضين الخامس $(\beta_k \neq 0)$ والسادس $(\beta_k = \beta)$, ولذلك يجب علينا أن نقوم بإيجاد معاملات العلاقات الخطية بين X Y في جميع المجتمعات $X_k = a_k + \beta_k Y_k$ بطريقة المربعات الصغرى، ثم القيام باختبار عدم وجود فروقات معنوية بين المعاملات β_k , وذلك بوضع الفرضيتين كما يلي:

$$H_0 : \beta_k = \beta_0 \qquad : k = 1\ 2\ 3 \dots g$$

$H_1 : \beta_k \neq \beta_0$ على الأقل من أجل قيمة واحدة لـ k

حيث أن β_0 هي قيمة افتراضية يضعها الباحث أو يحسبها من متوسط المعاملات β_k، وللتحقق من H_0 نقوم بحساب قيمة اختبار (ستودينت) t المعرف بالعلاقة:

$$t_k = \frac{\beta_k - \beta_0}{S(\beta_k)} \qquad : k = 1\ 2\ 3 \dots g \qquad (83-7)$$

حيث $S(\beta_k)$ هو الانحراف المعياري للمعامل β_k .

ثم نقوم بمقارنة قيمة t_k المحسوبة مع قيمة t الحرجة والمقابلة لـ $(n-1)$ درجة حرية ولنصف مستوى الدلالة $\left(\frac{\propto}{2}\right)$.

فإذا كانت $|t_k| \leq t_{n-1}\left(\frac{\propto}{2}\right)$ فإننا نقبل H_0 ونعتبر المعاملات β_k متساوية باحتمال 0.95، وبذلك يكون الافتراض السادس محققاً.

ولاختبار تحقق الافتراض الخامس نطبق الاختبار t على الفرضيتين التاليتين كما يلي:

$$H_0 : \beta_0 = 0 \qquad H_1 : \beta_0 \neq 0$$

ونطبق نفس المؤشر المعرف في (7−83) على جميع β_k بطريقة إعادة الاختبار. فإذا كانت النتيجة هي رفض H_0 من أجل جميع β_k، فإننا نعتبر أن $\beta_0 \neq 0$ وأنه يوجد علاقة خطية معنوية بين X وY في تلك المجتمعات، ويمكن التحقق من الافتراض الخامس عن طريق معامل الارتباط r أو معامل التحديد R^2 ...الخ.

وبناءً على ذلك يتم إجراء تحليل التغاير ANCOVA للتابع X المرتبط بمتحول إضافي Y. وذلك ضمن تحقق الافتراضات المذكورة أعلاه.

وبعد هذه المقدمة ننتقل إلى إجراء التحليل اللازم للنموذج (7−81) ونضع الفرضيتين الخاصتين به (العدم والبديلة) على الشكل التالي:

$$H_0 : \ltimes_1 = \ltimes_2 = \ltimes_3 = \cdots = \ltimes_g = 0$$

$$H_1 : \ltimes_k \neq 0 \quad \text{من أجل } k \text{ واحدة على الأقل} \qquad (84-7)$$

ولاختبار الفرضيات (7−84) علينا أن نقوم بحساب مجاميع مربعات الانحرافات لكل من المتحولين X وY ولجدائهما $X * Y$, وذلك باستخدام العلاقات الرياضية التالية (مع الانتباه هنا إلى أن r هو الحجم الموحد للعينات المسحوبة من تلك المجتمعات وg عدد المجتمعات):

بالنسبة للمتحول X نجد أن:

$$SST_x = \sum_{k=1}^{g}\sum_{i=1}^{r}(x_{ki} - \bar{X})^2 = \sum_{k=1}^{g}\sum_{i=1}^{r} x_{ki}^2 - \frac{X^2}{g * r} \qquad (85-7)$$

حيث أن: $X = \sum_{k=1}^{g}\sum_{i=1}^{r} x_{ki}$ هو مجموع قيم X في جميع العينات.

$$SSA_x = \sum_{k=1}^{g}(\bar{X}_k - \bar{X})^2 = \sum_{k=1}^{g}\frac{\mathrm{x}_k^2}{r} - \frac{X^2}{g * r} \qquad (86-7)$$

حيث أن: $X_k = \sum_{i=1}^{r} x_{ki}$ هو مجموع قيم X في العينة k فقط و$\bar{X}_k$ متوسطها.

$$SSE_x = \sum_{k=1}^{g}\sum_{i=1}^{r}(x_{ki} - \bar{X}_k)^2 = SST_x - SSA_x \qquad (87-7)$$

ولها درجات الحرية : $\quad g(r-1) \;\; = (gr-1) \;\; (g-1)$

أما بالنسبة للمتحول Y فنجد أن:

$$SST_y = \sum_{k=1}^{g}\sum_{i=1}^{r}(y_{ki} - \bar{Y}) = \sum_{k=1}^{g}\sum_{i=1}^{r} y_{ki}^2 - \frac{Y^2}{g * r} \qquad (88-7)$$

حيث أن: $Y = \sum\sum y_{ki}$ هو مجموع قيم Y في جميع العينات .

$$SSA_y = \sum_{k=1}^{g}(\bar{Y}_k - \bar{Y})^2 = \sum_{k=1}^{g}\frac{Y_k^2}{r} - \frac{Y^2}{g * r} \qquad (89-7)$$

حيث أن: $Y_k = \sum_{i=1}^{r} y_{ki}$ هو مجموع قيم Y في العينة k فقط ومتوسطها $\bar{Y}_k$.

$$SSE_y = \sum_{k=1}^{g}\sum_{i=1}^{r}(y_{ki} - \bar{Y}_k)^2 = SST_y - SSA_y \qquad (90-7)$$

أما بالنسبة لمجاميع الجداء $X * Y$ فنجد أن:

$$SPT_{xy} = \sum_{k=1}^{g}\sum_{i=1}^{r}(x_{ki} - \bar{X})(y_{ki} - \bar{Y}) = \sum_{k=1}^{g}\sum_{i=1}^{r} x_{ki} * y_{ki} - \frac{X * Y}{gr} \qquad (91-7)$$

حيث أن: $X = \sum\sum x_{ki}$ وأن $Y = \sum\sum y_{ki}$ وأن درجة حريته $(gr-1)$.

$$SPE_{xy} = \sum_{k=1}^{g}\sum_{i=1}^{r}(x_{ki} - \bar{X}_k)(y_{ki} - \bar{Y}_k) = \sum_{k=1}^{g}\left[\sum_{i=1}^{r}(x_{ki} * y_{ki}) - \frac{X_k * Y_k}{gr}\right] = \sum_{k=1}^{g} SPE_k \quad \mathbf{(92 - 7)}$$

حيث أن: $X_k = \sum_{i=1}^{r} x_{ki}$ وأن: $Y_k = \sum_{i=1}^{r} y_{ki}$ وأن درجة حريته $= g(r-1)$.

$$SPA_{xy} = \sum_{k=1}^{g}(\bar{X}_k - \bar{X})(\bar{Y}_k - \bar{Y}) = \sum_{k=1}^{g}\frac{X_k Y_k}{r} - \frac{XY}{gr} = SPT_{xy} - SPE_{xy} \qquad \mathbf{(93 - 7)}$$

وهنا نشير إلى أن المجاميع الجدائية SPT_{xy} وSPE_{xy} SPA_{xy} يمكن أن تكون سالبة على عكس مجاميع المربعات لـ X وY الموجبة دائماً , ولسهولة العرض نضع هذه المربعات والجداءات في جدول منظم كالتالي:

جدول (7–17): جدول التحليل قبل التعديل

مصدر التباين	درجة الحرية dk	مربعات المتحول X	مربعات المتحول Y	مجاميع الجداء $X * Y$
المجتمعات (المعالجات)ِ SSA	$g-1$	SSA_x	SSA_y	SPA_{xy}
الخطأ العشوائي SSE	$g(r-1)$	SSE_x	SSE_y	SPE_{xy}
الاجمالي SST	$g\,r-1$	SST_x	SST_y	SPT_{xy}

وأخيراً نقوم بحساب مجاميع مربعات الانحرافات المعدلة بعد عزل تأثير المتحول الإضافي Y، من العلاقات التالية:

$$(SST)' = (SST)_x - \frac{(SPT)_{xy}^2}{(SST)_y} \quad : d\!f = gr - 2 \qquad (94 - 7)$$

$$(SSE)' = (SSE)_x - \frac{(SPE)_{xy}^2}{(SSE)_y} \quad : d\!f = g(r-1) - 1 \qquad (95 - 7)$$

$$(SSA)' = (SST)' - (SSE)' \quad : d\!f = g - 1 \qquad (96 - 7)$$

وبناء على ذلك يمكننا أن نضع هذه النتائج في جدول تحليل التباين المشترك المعدل والذي يأخذ الشكل(7–18) التالي، ومنه نلاحظ أنه لقد تم تخفيض درجة حرية الاجمالي ودرجة حرية الخطأ التجريبي بمقدار درجة واحدة عما كانت عليه في الجدول (7–17)، وذلك لأننا استخدمنا بيانات العينة في تقدير β والذي يساوي: $\tilde{\beta} = \frac{(SSE)_{xy}^2}{SSE_y}$.

أما المقدار: $\frac{(SPT)_{xy}^2}{SST_y}$ فهو مجموع مربعات الانحدار للنموذج (7–81) بدون اعتبار تأثير المجتمعات. والذي يأخذ الشكل التالي (بدون $\propto_k$) :

$$x_{ki} = \mu_x + \beta(y_{ki} - \bar{Y}) + e_{ki}$$

وهو يعبر عن كمية التباين في قيمة X الناجمة عن المتحول المستقل Y .

جدول (7–18): جدول تحليل التباين المشترك المعدل:جدول ANCOVA

مصدر التباين	الرمز	درجة حرية	متوسط مجموع المربعات المعدلة	قيمة F المحسوبة
بين المجتمعات أو المعالجات	$(SSA)'$	$g-1$	$MSA' = \frac{SSA'}{g-1}$	$F = \frac{MSA'}{MSE'}$
الأخطاء العشوائية	$(SSE)'$	$g(r-1)-1$	$MSE' = \frac{SSE'}{g(r-1)-1}$	———
الاجمالي (المعدل)	$(SST)'$	$g*r-2$	———	———
	حيث أن : r هو حجم العينة الموحد في المجتمعات			

ثم نقوم بإجراء الاختبار اللازم باستخدام المؤشر F المعرف بالعلاقة التالية :

$$F = \frac{\frac{SSA'}{g-1}}{\frac{SSE'}{(g(r-1)-1)}} = \frac{MSA'}{MSE'} \sim F_{(g-1),(g(r-1)-1)} \qquad (97-7)$$

وهو يخضع لتوزيع F بدرجتي حرية $v_1 = (g-1)$ و $v_2 = (g(r-1)-1)$، لذلك نقارن هذه القيمة المحسوبة مع القيمة الحرجة للمتحول F عند مستوى دلالة ∝ ونتخذ القرار كما يلي:

إذا كانت $F \le F_{v_1,v_2}(\propto)$ نقبل فرضية العدم H_0

أما إذا كانت $F > F_{v_1,v_2}(\propto)$ نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1 بمستوى دلالة ∝ .

ملاحظة: لا يجوز إجراء تحليل التباين المشترك ANCOVA , قبل حساب معادلة الانحدار الخطي بين X وY واختبار معنوية β , والتي سيكون لها الصيغة التالية :

$$X = a + \beta Y \qquad (98-7)$$

فإذا كانت قيمة β غير معنوية (معدومة $\beta = 0$ في جميع المجتمعات) فإن ذلك يعني أن X غير مرتبط بـ Y ، ولا داعي لإدخال Y في النموذج وتعديله، وعندها نقوم بحساب تحليل التباين البسيط باتجاه واحد كالعادة على X فقط .

أما إذا كانت قيمة β معنوية (غير معدومة), فإننا نختبر قيمها إذا كانت متساوية في جميع المجتمعات أم لا.

فإذا كانت قيم β متساوية في جميع المجتمعات نقوم بإجراء تحليل ANCOVA ونعدل النموذج كما سبق.

أما إذا كانت قيم β غير متساوية في المجتمعات المدروسة, فإننا نقوم بحساب معادلات الانحدار في كل مجتمع على حدة ونتخذ القرار المناسب حول الأسلوب المناسب لتفسير أسباب تغيرات X .

ملاحظة 2: يمكن تعميم أسلوب تحليل ANCOVA على متحولين إضافيين أو أكثر Y_1 و Y_2 مرتبطين بالمتحول التابع المدروس X، وذلك باتباع نفس المعالجة وباستخدام نفس الفرضيات والاختبارات.

مثال (7–6): لنفترض أننا نريد دراسة تغيرات علامة الرياضيات للطلاب الدارسين في 3 مدارس محددة (3 مجتمعات) . لذلك نرمز لعلاماتهم في مقرر الرياضيات فيها بـ X، ونريد أن ندرس تغيرات X الناتجة على اختلاف تلك المدارس(أي دراسة تأثير المدارس على علامة الرياضيات). إلا أن أحد المختصين لفت انتباهنا إلى أن علامة الرياضيات X مرتبطة أيضاً بدرجات ذكاء الطالب Y , لذلك يجب عزل تأثيره. ولهذا علينا استخدام أسلوب تحليل ANCOVA وفق النموذج المعدل (7–81) التالي:

$$x_{ki} = \mu_x + \propto_k + \beta(y_{ki} - \mu_y) + e_{ki} \qquad (99-7)$$

ولذلك قمنا بسحب (3) عينات عشوائية من طلاب هذه المدارس بحجم موحد $n_k = r = 10$ ، وأخذنا منهم علامات الرياضيات وحددنا درجة الذكاء لكل طالب، فكانت كما في الجدول التالي (العلامات X والدرجة Y حسبت من 100).

جدول (7–19): العلامات X والدرجات Y مع مجاميعهما

رقم الطالب	المدرسة الاولى		المدرسة الثانية		المدرسة الثالثة		المجموع	
	Y_{1i}	X_{1i}	Y_{2i}	X_{2i}	Y_{3i}	X_{3i}	Y	X
1	73	55	50	76	82	62		
2	60	70	66	80	88	90		
3	45	30	90	86	90	82		
4	33	27	86	70	50	40		
5	90	89	91	85	70	42		
6	68	50	80	73	75	80		
7	77	60	50	40	80	90		
8	80	98	40	35	95	60		
9	85	79	47	25	40	30		
10	70	82	90	60	50	44		
$\sum$ المجموع	680	640	690	630	720	620	$Y = 2090$	$X = 1890$
معادلات الانحدار	$\tilde{X}_1 = -12.87 + 1.130Y$ $r_1 = 0.836$		$\tilde{X}_2 = 10.42 + 0.762Y_2$ $r_2 = 0.712$		$\tilde{X}_3 = -3.54 + 0.910Y_3$ $r_3 = 0.7745$			

ولدراسة تغيرات X بمعزل عن Y نقوم أولاً بحساب معادلات الانحدار الخطي لعلاقة X بـ Y في كل مدرسة على حدة فنجد أنها تساوي (انظر السطر الأخير من الجدول):

$$\tilde{X}_1 = a + bY_1 = -12.87 + 1.130\, Y_1 \qquad (r_1 = 0.836)$$
$$\tilde{X}_2 = a + bY_2 = 10.42 + 0.762\, Y_2 \qquad (r_2 = 0.712)$$
$$\tilde{X}_3 = a + bY_3 = -3.54 + 0.910\, Y_3 \qquad (r_3 = 0.7745)$$

وبدراسة هذه المعادلات ومعاملات الارتباط فيها نلاحظ أن قيم b فيها ليست معدومة (وليست قريبة من الصفر) . ولذلك فإننا نتابع التحليل ANCOVA دون الخوض في مسألة البرهان على ذلك. ونترك مسألة البرهان على تساوي أو عدم تساوي قيم b في المجتمعات الثلاثة إلى القارئ على سبيل التمرين .

ولمتابعة تحليل ANCOVA نضع النموذج الرياضي كما يلي:

$$x_{ki} = \mu_x + \propto_k + \beta(y_{ki} - \mu_y) + e_{ki}$$

ثم نضع الفرضيتين كما يلي:

$$H_0 : \propto_1 = \propto_2 = \propto_3 = \cdots = \propto_g = 0 \qquad \left(\text{بشرط } \sum \propto_k = 0\right)$$

$H_1 : \propto_k \neq 0$ من أجل k واحدة على الأقل

وحتى نستطيع حساب مجاميع المربعات والجداءات السابقة، قمنا بحساب مجاميع قيم X ومجاميع قيم Y الكلية والهامشية، ووضعناها في الجدول (7−19) السابق, كما قمنا بحساب مجاميع المربعات والجداءات لـ X و Y ووضعناها في الجدول (7−20) , وبناء على معطيات هذين الجدولين نجد أن:

جدول (7−20): جدول مساعد لحساب مربعات وجداءات Y و X من عناصر العينات

الرموز	المدرسة الأولى			المدرسة الثانية			المدرسة الثالثة			المجاميع		
	y_{1i}^2	x_{1i}^2	$y_{1i}x_{1i}$	y_{2i}^2	x_{2i}^2	$y_{2i}x_{2i}$	y_{3i}^2	x_{3i}^2	$y_{3i}x_{3i}$	Y^2	X^2	Y*X
1	5184	3025	3960	2500	5776	3800	6721	3844	5084			
2	3600	4900	4200	4356	6400	5280	7744	5100	7920			
3	2025	900	1350	8100	7396	7740	8100	6724	7380			
4	1089	729	891	7596	4900	6020	2500	1600	2000			
5	5100	7921	8010	8281	7225	7735	4900	1764	2940			
6	4624	2500	3400	6400	5329	5840	5625	6400	6000			
7	5929	3600	4620	2500	1600	2000	6400	8100	7200			
8	6400	9604	7840	1600	1225	1400	9025	3600	5700			
9	7225	6241	6715	2209	625	1175	1600	900	1200			
10	4900	6724	5740	8100	3600	5400	2500	1936	2200			
المجموع $\sum_{i=1}^{10}$	49067	46144	46726	51442	44067	46390	55118	72968	47624	155636	133188	140740

الرموز	المدرسة الأولى			المدرسة الثانية			المدرسة الثالثة			المجاميع		
	y_{1i}^2	x_{1i}^2	$y_{1i}x_{1i}$	y_{2i}^2	x_{2i}^2	$y_{2i}x_{2i}$	y_{3i}^2	x_{3i}^2	$y_{3i}x_{3i}$	Y^2	X^2	Y*X
متوسط [1](*) مربعات المجاميع	46240	40960	ــــ	47610	39690	ــــ	51840	38440	ــــ	ــــ	ــــ	ــــ
متوسط (**) جداء المجاميع	ــــ	ــــ	43520	ــــ	ــــ	43470	ــــ	ــــ	44640	ــــ	ــــ	ــــ
$(SSE)_k$	2836	5184	ــــ	3832	4386	ــــ	3278	4528	ــــ	9946	14098	ــــ
$(SPE)_k$	ــــ	ــــ	3206	ــــ	ــــ	2920	ــــ	ــــ	2984	ــــ	ــــ	9110

$$Y_1 = \sum_{i=1}^{10} y_{1i} = 680 \qquad X_1 = \sum_{i=1}^{10} x_{1i} = 640$$

$$Y_2 = \sum_{i=1}^{10} y_{2i} = 690 \qquad X_2 = \sum_{i=1}^{10} x_{2i} = 630$$

$$Y_3 = \sum_{i=1}^{10} y_{3i} = 720 \qquad X_3 = \sum_{i=1}^{10} x_{3i} = 620$$

$$Y = \sum_{k=1}^{3}\sum_{i=1}^{10} y_{ki} = 2090 \qquad X = \sum_{k=1}^{3}\sum_{i=1}^{10} x_{ki} = 1890$$

ومن الجدول (7−19) نجد أن:

$$T_y = \sum\sum y_{ki}^2 = 155636 \quad ,T_x = \sum\sum x_{ki}^2 = 133188$$

وبالاعتماد على العلاقات من (7−84) إلى(7−88) السابقة وعلى الجدولين (7−19) و(7−20) , قمنا بحساب مجاميع المربعات والجداءات لكل من Y وX كما يلي:

$$(SST)_y = T_y - \frac{Y^2}{gr} = 155636 - \frac{(2090)^2}{30} = 10032.67$$

$$(SSE)_y = \sum_{k=1}^{3}(SSE_k)_y = 2836 + 3832 + 3278 = 9946$$

$$(SSA)_y = (SST)_y - (SSE)_y = 10032.67 - 9946 = 86.67$$

(*) تم حساب متوسط مربعات المجاميع اعتماداً على الجدول (4−19) ومن العلاقة: $\frac{Y_k^2}{r}$ ثم $\frac{X_k^2}{r}$ مثل: $\frac{Y_1^2}{r} = \frac{(680)^2}{10} = 46240$

(**) كما تم حساب متوسط جداءات المجاميع اعتماداً على الجدول (5−19) ومن العلاقة : $\frac{Y_kX_k}{r}$ مثل: $\frac{Y_1X_1}{r} = \frac{(680)(640)}{10} = 43520$

$$(SST)_x = T_x - \frac{X^2}{gr} = 133188 - \frac{(1890)^2}{30} = 14118$$

$$(SSE)_x = \sum_{k=1}^{3} (SSE_k)_x = 5184 + 4386 + 4528 = 14098$$

$$(SSA)_x = (SST)_x - (SSE)_x = 14118 - 14098 = 20$$

$$(SPT)_{xy} = T_{xy} - \frac{X * Y}{gr} = 140740 - \frac{(2090)(1890)}{30} = 74905$$

$$(SPE)_{xy} = \sum_{k=1}^{3} (SSE_k)_{xy} = 3206 + 2920 + 2984 = 9110$$

$$(SPA)_{xy} = (SPT)_{xy} - (SPE)_{xy} = 74905 - 9110 = 65795$$

ولتسهيل الإجراءات العملية نقوم بوضع نتائج هذه الحسابات في جدول مختصر كما يلي:

جدول (7–21): مربعات وجداءات X و Y

مصدر التباين	درجة الحرية	مجموع مربعات X	مجموع مربعات Y	مجموع الجداءات $X * Y$
المجتمعات (المعالجات) SSA	2	20	86.67	65795
الأخطاء (البواقي) SSE	27	14098	9946	9110
الاجمالي SST	29	14118	10032.67	74905

ثم نقوم بحساب المجاميع المعدلة للتابع X من العلاقات التالية :

$$(SST)' = (SST)_x - \frac{(SPT)_{xy}^2}{(SST)_y} = 14118 - \frac{(74905)^2}{10032.67} = 8525.5$$

$$(SSE)' = (SSE)_x - \frac{(SPE)_{xy}^2}{(SSE)_y} = 14098 - \frac{(9110)^2}{9946} = 5753.73$$

$$(SSA)' = (SST)' - (SSE)' = 8525.5 - 5753.73 = 2771.77$$

ثم نقوم بوضع المجاميع المعدلة الأخيرة في جدول تحليل ANCOVA فنحصل على أن:

جدول (7–22): جدول تحليل التغاير ANCOVA حيث أن: $g = 3$ و $r = 10$

مصدر التباين	درجة الحرية	مجاميع المربعات المعدلة	متوسط مجاميع المربعات المعدلة	قيمة F المحسوبة
المجتمعات أو المعالجات SSA	$g - 1 = 2$	2771.77	1385.885	$F = \frac{1385.885}{221.30} = 6.25$
الأخطاء العشوائية SSE'	$g(r-1) - 1 = 26$	5753.73	221.30	———
Totel الاجمالي	$gr - 2 = 28$	8525.5	———	———

ثم نقوم بحساب قيمة مؤشر الاختبار F من العلاقة:

$$F = \frac{MSSA'}{MSSE'} = \frac{1385.885}{221.30} = 6.25$$

ثم نبحث في جداول F عن القيمة الحرجة لمتحول F المقابلة لمستوى الدلالة ($\propto = 0.05$) ولدرجتي الحرية $v_1 = 2$ و$v_2 = 26$ فنجد أن:

$$F_{v_1,v_2}(\propto) = F_{2\,,26}(0.05) = 3.37$$

وبالمقارنة نجد أن: $(F = 6.25) > 3.37$، لذلك نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1 ، التي تقول أنه يوجد تأثير للمجتمعات المدروسة على علامات الطلاب في مقرر الرياضيات، وذلك بعد عزل أثر المتحول Y المرتبط مع المتحول X والمتمثل بدرجة الذكاء لهؤلاء الطلاب.

ملاحظة: يستخدم تحليل التباين المشترك ANCOVA لتحقيق هدفين هما:

1- **لزيادة دقة التجربة** وذلك لأنه يعمل على عزل وإبعاد المتحول الإضافي Y المرتبط مع X. وحتى نظهر للقارئ معنى هذا الكلام ننشأ جدول تحليل التباين البسيط ANOVA قبل استبعاد المتحول Y وتعديل المجاميع . فنجد أنه كما يلي:

جدول (7-23): جدول ANOVA **قبل التعديل**

مصدر التباين	درجة الحرية	مجاميع مربعات الانحرافات		
		المربعات X	المتوسطات XY	F
المجتمعات (SSA)	2	20	10	$F = 0.019$
الأخطاء (SSE)	27	14096	522.07	
الإجمالي (SST)	29	14118		

ومنه نحسب قيمة المؤشر F للمتحول X فنجد أن:

$$F_x = \frac{20/2}{14098/27} = \frac{10}{522.15} = 0.019$$

وبمقارنة هذه القيمة المحسوبة مع القيمة الحرجة $F_{2\,,27}(0.05) = 3.37$ نجد أن $F < 3.37$, لذلك كان يمكن أن نقبل H_0 التي تقول أن $\propto_k = 0$، ونقول أن المجتمعات المدروسة لا تؤثر على علامات الرياضيات. وهكذا نجد أن هذه النتيجة تخالف النتيجة السابقة التي حصلنا عليها بعد عزل تأثير درجة الذكاء Y. وهكذا نجد أن تحليل التباين المشترك ANCOVA يكون مفيداً في تدقيق النتائج وتصويبها.

2- **لتقليل نسبة الخطأ وزيادة الكفاءة النسبية:** ففي مثالنا الحالي نجد أن نسبة الخطأ (من المجموع) قبل التعديل كانت تساوي: $P_1 = \frac{SSE_X}{SST_X} = \frac{14098}{14118} = 0.9986$

أما نسبة الخطأ بعد التعديل (من المجموع المعدل) فتساوي:

$$P_2 = \frac{SSE'}{SST'} = \frac{5753.73}{8525.5} = 0.6749$$

ولتقدير الكفاءة النسبية لتحليل التغاير نقوم بحسابها من العلاقة التالية:

$$RE = \frac{MSSE}{MSSE'} = \frac{SSE/g(r-1)}{SSE'/(g(r-1)-1)} = \frac{1}{1-r^2}$$

$$RE = \frac{14098/27}{5758.73/26} = \frac{523.14}{221.30} = 2.36$$

أي أن تطبيق تحليل التباين المشترك ANCOVA قد أدى إلى زيادة كفاءة التحليل ودقة التجربة بمقدار 2.36 مرة.

ملاحظة: يمكن التوسع في تطبيقات تحليل التباين المشترك لزيادة حساسية التجربة، وذلك باستخدام معيارين وإجراء التحليل باتجاهين للتقليل من أثر المتحول الإضافي Y، وعندها نقوم بنفس الخطوات المذكورة في هذه الفقرة مع إجراء بعض التعديلات اللازمة عليها . وسنترك ذلك للقارئ للتعمق بها.

الفصل الثامن
الارتباط والانحدار والسلاسل الزمنية

8-0: تمهيد :

يتناول هذا الفصل دراسة العلاقة الارتباطية بين متحولين متفاعلين , بحيث يكون أحدهما المتحول المؤثر (المستقل) ونرمز له بـ x والآخر المتحول المتأثر (التابع أو الدالة) ونرمز له بـ y . وبصورة عامة يمكننا عند دراسة العلاقات الارتباطية بين الظواهر تقسيم الظواهر المدروسة في لحظة ما أو في مكان ما إلى صنفين أساسيين هما:

ظواهر مسببة: وهي جملة العوامل التي يمكن أن تؤثر على الظواهر المستهدفة بالدراسة.

ظواهر ناتجة: وهي جملة الظواهر المستهدفة بالدراسة والمتأثرة بالعوامل السابقة.

وهنا نشير إلى أن عملية البحث عن العلاقة الارتباطية بين ظاهرتين أو أكثر يجب أن تتوفر فيها الشروط التالية:

1- أن تكون بين الظاهرتين علاقة جدلية وسببية واضحة مثل: علاقة التدخين بسرطان الرئة أو بلون الأسنان ,علاقة الغيوم بالمطر..... الخ.

2- أن تكون إحدى الظاهرتين مسببة والأخرى ناتجة، ويجب تحديد طبيعة كل منهما مسبقاً.

3- أن تكون الظاهرتان قابلتين للقياس حسب واحدة قياس معينة لكل منهما أو قابلتين للتصنيف أو الترتيب أو التبويب حسب مؤشر معين لكل منهما.

4- أن تكون القياسات المأخوذة أو التصنيفات والتبويبات المتبعة متقابلة من حيث المكان أو الزمان أو كلاهما معاً.

وأخيراً نشير إلى أنه يمكن تصنيف العلاقات بين المتحولات والظواهر إلى نوعين أساسيين هما:

العلاقات التابعية: وهي العلاقات التي تكون محددة بقوانين رياضية دقيقة مثل : مساحة المستطيل = الطول x العرض.

العلاقات الارتباطية: وهي العلاقات التي يتم التعبير عنها بواسطة معادلات رياضية تقريبية (غير دقيقة تماماً) مثل محيط الدائرة .وكذلك مثل: علاقة الوزن x بالطول y للمواليد، أو مثل: علاقة الدخل y بالعمر x . أو علاقة عمر الزوج بعمر الزوجة .....الخ.

1-8: طرائق الكشف عن العلاقة الارتباطية :

قبل أن نتعرض لطرائق الكشف عن العلاقة الارتباطية نفترض أننا نريد دراسة العلاقة بين متحولين كميين X و Y . وأننا حصلنا على n قياساً من القياسات المتقابلة عنها , فكانت كما يلي:

$$X: x_1\ , x_2\ , x_3\ , x_4 \ldots\ x_i \ldots \ldots \ldots\ x_n \qquad (1-8)$$
$$Y: y_1\ , y_2\ , y_3\ , y_4 \ldots\ y_i \ldots \ldots \ldots\ y_n$$

وتسمي هذه القياسات أو القيم المتقابلة بالسلسلة الارتباطية, كما نفترض أننا قمنا بحساب المتوسط الحسابي والتباين لكل منهما من العلاقات التالية:

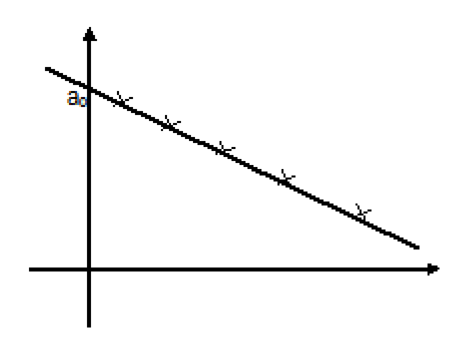
$$\bar{x} = \frac{1}{n} \sum^n x_i \qquad \bar{y} = \frac{1}{n} \sum^n y_i \qquad (2-8)$$

$$\sigma_x^2 = \frac{1}{n} \sum^n (x_i - \bar{x})^2 \qquad \sigma_y^2 = \frac{1}{n} \sum^n (y_i - \bar{y})^2 \qquad (3-8)$$

أما أهم طرائق الكشف عن وجود العلاقة الارتباطية بين Y و X فهي:

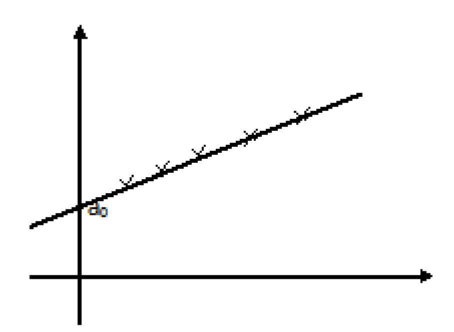
(8-1-1) – طريقة شكل الانتشار:

وهي تعتمد على رسم الشكل البياني للنقاط الهندسية (x_i , y_i) على المحورين الإحداثيين (وهي عبارة عن القياسات المتقابلة للمتحولين ((x , y)) ونضع المتحول المسبب (المستقل) x على المحور الأفقي OX والمتحول التابع y على المحور الشاقولي OY ، ثم نختار واحدات رسم وقياس مناسبة، ونتيجة لذلك نحصل على أحد الأشكال التالية:



X شكل (8-2): علاقة عكسية وخطية

$y = a_o + a\ X \quad : \quad a < o$

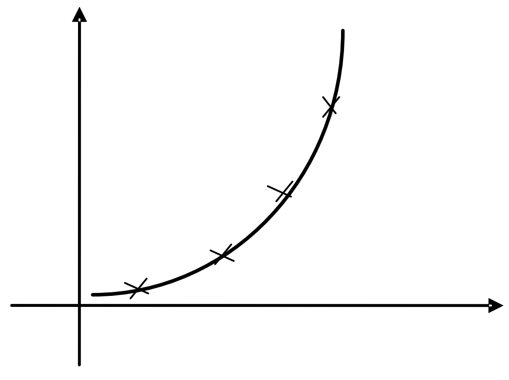


X شكل (8-1): علاقة طردية وخطية

$y = a_o + a\ X \quad : \quad a > o$

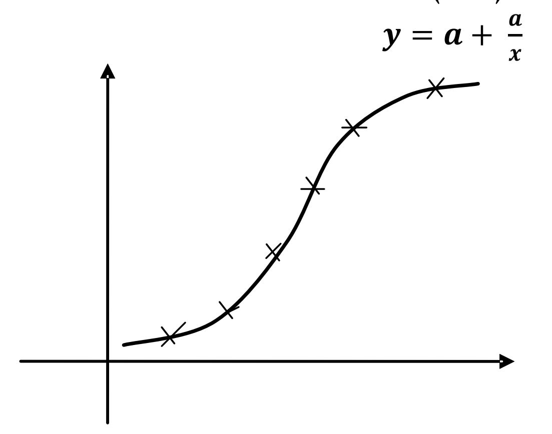
شكل (8–4): علاقة منحنية عكسية

$y = a + \frac{a}{x}$

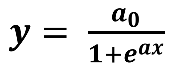


شكل (8–3): علاقة منحنية أسية

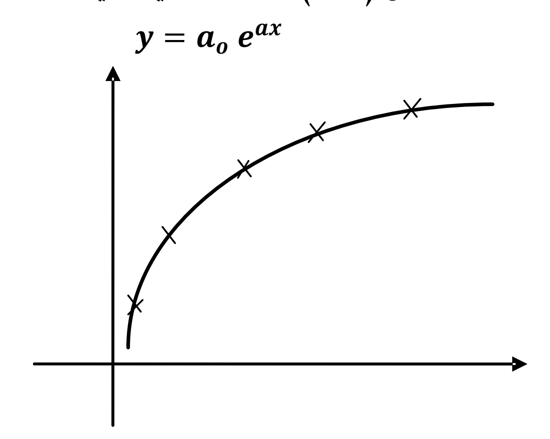
$y = a_o\, e^{ax}$



شكل (8– 6): علاقة منحنية لوجستية (منطقية)

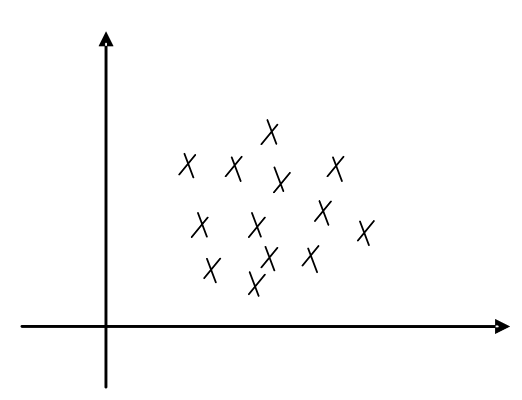


$y = \frac{a_0}{1+e^{ax}}$



شكل (8– 5): علاقة منحنية لوغاريتمية

$y = a_o + a \log x$



شكل (8–7): عدم وجود علاقة واضحة

من هذه الأشكال نلاحظ أن العلاقة بين المتحولين Y و X يمكن:

- أن تكون خطية (على شكل مستقيم) طردية أو عكسية (موجبة أو سالبة).
- أن تكون منحنية طردية (من الدرجة الثانية أو أسية) أو عكسية (بمقلوب x).
- أن تكون منحنية طردية (لوغاريتمية) أو منطقية (لوجستية).

والشكل الأخير (8–7) يشير بوضوح على عدم وجود علاقة بين المتحولين y و x .

ويكون الارتباط قوياً كلما كانت النقاط متجمعة باتجاه معين.

ويكون ضعيفاً كلما كانت النقاط مبعثرة على شكل الانتشار.

(8–1–2): طريقة المقارنة:

وتعتمد هذه الطريقة على ترتيب القيم المتقابلة (x_i , y_i) تصاعدياً حسب قيم المتحول x ثم القيام بدراسة تغيرات المتحول y مع تغيرات المتحول x ، ومن خلال ذلك يمكننا أن نلاحظ مباشرة فيما إذا كانت قيم y تتزايد أو تتناقص مع تزايد قيم x (بغض النظر عن بعض القيم الشاذة) وعندها نستنتج أن y مرتبط طردياً أو عكسياً مع x .

أما إذا كانت تغيرات المتحول y غير منتظمة ولا تتوافق لا طردياً ولا عكسياً مع تزايد قيم x فإننا نستنتج أن العلاقة بينهما تكون ضعيفة أو معدومة.

(8–1–3): طريقة معامل الارتباط الخطي Coefficient of linear correlation:

ويسمى هذا المعامل (**معامل بيرسون Pearson's Coefficient**).

ويستخدم هذا المعامل للكشف عن متانة العلاقة الارتباطية بين متحولين كميين y و x .

وبالاعتماد على الرموز الواردة في العلاقات (8–1) و (8–2) و (8–3) السابقة يُعرف هذا المعامل r بالعلاقة الأساسية التالية:

$$r = \frac{\sum_{i=1}^{n} [(x_i - \bar{x})\ (y_i - \bar{y})]}{n\ .\ \sigma_x\ .\ \sigma_y} \qquad (4-8)$$

وباستبدال كل من σ_x و σ_y من العلاقة (8–3) وإجراء بعض الاختصارات نحصل على العلاقة المكافئة لـ (8–4) التالية:

$$r = \frac{\sum^{n} [(x_i - \bar{x})\ (y_i - \bar{y})]}{\sqrt{\sum^{n} (x_i - \bar{x})^2}\ .\ \sqrt{\sum^{n} (y_i - \bar{y})^2}} \qquad (5-8)$$

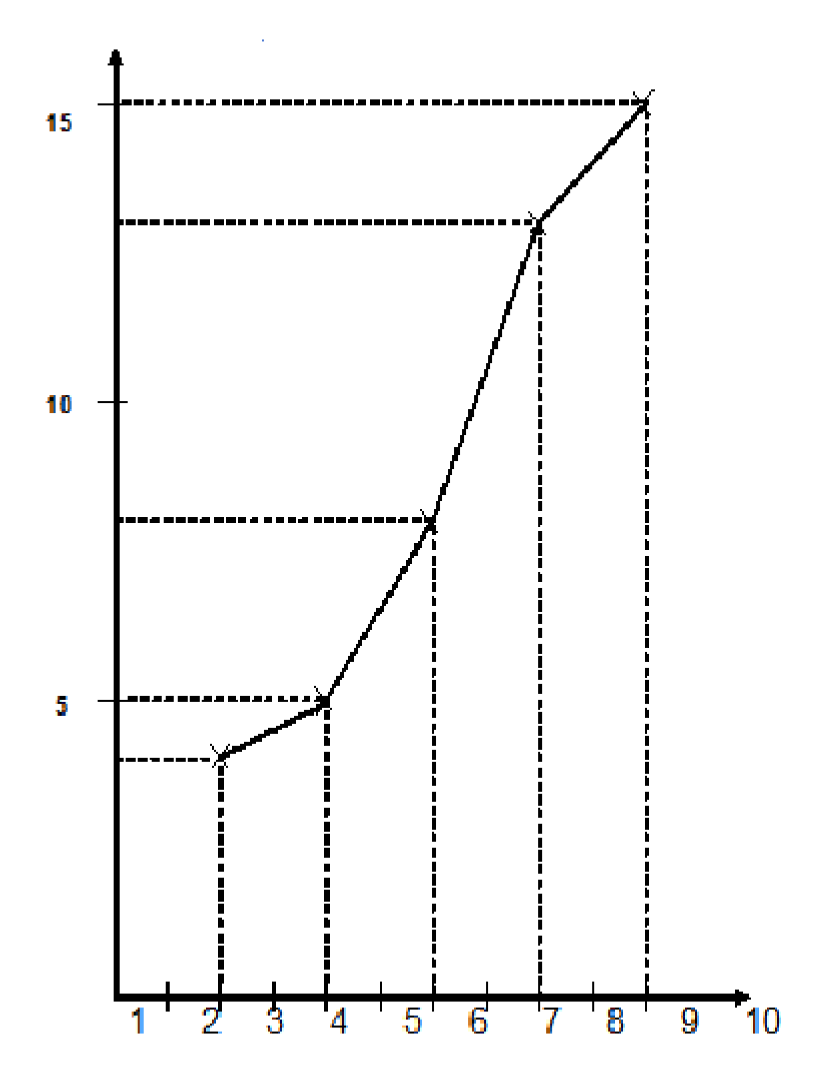
وبإجراء بعض العمليات الرياضية على البسط والمقام يمكننا اشتقاق عدة علاقات مكافئة للعلاقة (8–4) . وإن أسهل هذه العلاقات هي التي تأخذ الشكل التالي:

$$r = \frac{n \sum_{i=1}^{n} (x_i . y_i) - (\sum x_i)\ (\sum y_i)}{\sqrt{n \sum x_i^2 - (\sum x_i)^2}\ .\ \sqrt{n \sum y_i^2 - (\sum y_i)^2}} \qquad (6-8)$$

مثال (8–1) : ارسم الشكل البياني (شكل الانتشار) ثم احسب معامل الارتباط الخطي r لعلاقة دخل الطبيب y مع عدد الزوار x من القيم المتقابلة لكل منهما خلال خمسة أيام والتي كانت كما يلي:

رقم اليوم	1	2	3	4	5
$x : x_i$ عدد الزوار	2	4	6	8	10
$y : y_i$ قيمة الدخل (ألف)	4	5	8	13	15

<u>الحل</u>: إن شكل الانتشار لهذه القيم المتقابلة يأخذ الشكل البياني التالي:



الشكل (8-8): شكل الانتشار

إن شكل الانتشار يشير إلى وجود ارتباط خطي موجب بين المتحولين y و x .

ولحساب معامل الارتباط الخطي نعد الجدول المساعد التالي:

I	x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1	2	4	−4	−5	20	16	25
2	4	5	−2	−4	8	4	16
3	6	8	0	−1	0	0	1
4	8	13	2	4	8	4	16
5	10	15	4	6	24	16	36
Σ المجموع	30	45	0	0	60	40	94
النتائج	$\bar{x} = 6$	$\bar{y} = 9$				$\sigma_x^2 = 8$	$\sigma_y^2 = 18.8$

وبتعويض البيانات النهائية لهذا الجدول في العلاقة (8-4) نجد أن:

$$r = \frac{\sum^5 [(x_i - \bar{x})\,(y_i - \bar{y})]}{n\,.\,\sigma_x\,.\,\sigma_y} = \frac{60}{5\,.\,\sqrt{8}\,.\,\sqrt{18.8}} = 0.978$$

وهي قيمة كبيرة وتدل على وجود ارتباط متين بين y و x وهو ارتباط طردي (موجب).

وكان يمكن تعويض بيانات مجاميع الجدول السابق في العلاقة (8− 6) فنجد أن :

$$r = \frac{\sum^5 \ [(x_i - \bar{x})\ (y_i - \bar{y})]}{\sqrt{\sum\ (x_i - \bar{x})^2}\ .\ \sqrt{\sum\ (y_i - \bar{y})^2}} = \frac{60}{\sqrt{40}\ .\ \sqrt{94}} = 0.978$$

8−1−3−1: بعض خواص معامل الارتباط الخطي:

1− إن قيمة معامل الارتباط r محصورة في المجال [−1 , +1].

أي أن: $-1 \leq r \leq +1$

2− إذا كانت قيمة r قريبة من (+1) فإنها تدل على ارتباط متين طردي (موجب).

وإذا كانت قيمة r قريبة من (−1) فإنها تدل على ارتباط متين عكسي (سالب).

وإذا كانت قيمة r قريبة من (0) فإنها تدل على عدم وجود ارتباط بين المتحولين.

وإذا كانت قيمة $|r|$ بالقيمة المطلقة أصغر من 0.70 فإنها تدل على أن الارتباط الخطي ضعيف بين المتحولين ولا يصح الاعتماد عليه في البحوث والدراسات العلمية, وهنا يجب البحث عن معادلات منحنية لتمثيل العلاقة بين المتحولين y و x .

3− إن قيمة معامل الارتباط لا تتأثر بواحدات القياس لأي من المتحولين y أو x . لذلك يمكننا استخدام أية وحدات قياس لهما عند حسابه.

4− إن قيمة معامل الارتباط لا تتغير بتبديل مواقع المتحولين y و x .

أي أن: $r_{xy} = r_{yx}$

5− إن قيمة معامل الارتباط تعبر عن متانة الارتباط الخطي فقط (أي الذي يكون على شكل مستقيم) ولا يصح تطبيقه على الأشكال المنحنية للعلاقات الارتباطية.

8−1−4: طريقة معامل الارتباط الرتبي Rank coefficient of correlation:

ويسمى هذا المعامل (**معامل سبيرمان Sperman's Coefficient**).

ويستخدم هذا المعامل للكشف عن متانة العلاقة بين متحولين نوعيين مرتبين أو قابلين للترتيب y و x ويعرف بالعلاقة التالية:

$$r_s = 1 - \frac{6\ \sum(k_i - p_i)^2}{n\ (n^2 - 1)} \qquad (7-8)$$

حيث أن k_i هي الرتب المتصاعدة لحالات المتحول x .

وأن p_i هي الرتب المتصاعدة لحالات المتحول y المقابلة لحالات x .

ولتطبيق هذا المعامل لا بد من إجراء ترتيب الحالات المتقابلة (x_i , y_i) حسب حالات x المتصاعدة ثم وضع الرتب المتسلسلة تصاعدياً لكل من x و y .

مثال (8−2) : في دراسة لعلاقة مستوى تعليم الزوج مع مستوى تعليم الزوجة في إحدى القرى سحبنا عينة عشوائية بحجم $n = 8$ أسر فحصلنا على المعلومات التالية :

رقم الأسرة	1	2	3	4	5	6	7	8
مستوى تعليم الزوج x_i	إعدادية	متعلم	أمي	ابتدائية	عليا	ثانوية	جامعي	متوسطة
مستوى تعليم الزوجة y_i	ابتدائية	متعلمة	أمية	متعلمة	جامعية	ثانوية	متوسطة	إعدادية

والمطلوب : حساب معامل الارتباط الرتبي r_s لعلاقة مستوى تعليم الزوج بمستوى تعليم الزوجة .

الحل : قبل حساب قيمة هذا المعامل نلاحظ أولاً أنه لا يمكننا حساب معامل الارتباط الخطي لأنه معامل خاص بالمتحولات الكمية ولا يمكن استخدامه في هذه الحالة . لذلك نقوم بما يلي :

نرتب معلومات الحالات المتقابلة السابقة حسب حالات x المتصاعدة فنحصل على الجدول التالي :

رقم الأسرة	1	2	3	4	5	6	7	8
حالات x المرتبة	أمي	متعلم	ابتدائية	إعدادية	ثانوية	متوسطة	جامعية	عليا
حالات y المقابلة	أمية	متعلمة	متعلمة	ابتدائية	ثانوية	إعدادية	متوسطة	جامعية
رتب حالات x k_i	1	2	3	4	5	6	7	8
رتب حالات y p_i	1	2.5	2.5	4	6	5	7	8
$k_i - p_i$	0	−0.5	+0.5	0	−1	+1	0	0
$(k_i - p_i)^2$	0	0.25	0.25	0	1	1	0	0

وبذلك نجد أن قيمة معامل الارتباط الرتبي تساوي :

$$r_s = 1 - \frac{6\,[0 + 2.5 + 2.5 + 0 + 1 + 1 + 0 + 0]}{8\,(8\ *\ 8 - 1)} = 1 - \frac{15}{504} = 0.97$$

وهذا يعني أن علاقة الارتباط بين مستويي التعليم للزوج والزوجة متينة جداً وهي علاقة طردية لأن قيمة r_s موجبة.

8-2: تمثيل العلاقة الارتباطية (الانحدار) Regression:

يمكن تمثيل العلاقة الارتباطية بين أي متحولين كميين y و x حسب شكل الانتشار بواسطة أحد النماذج الرياضية التالية:

1- معادلة الخط المستقيم والتي تأخذ الشكل التالي: $y = a + bx$

2- معادلة الدرجة الثانية والتي تأخذ الشكل التالي: $y = a + b_1x + b_2x^2$

3- معادلة النموذج الأسي والتي يكون لها الشكل التالي: $y = a\ e^{bx}$

4- معادلة النموذج العكسي والتي يكون لها الشكل التالي: $y = a + \frac{b}{x}$

5- معادلة الشكل اللوغاريتمي والتي يكون لها الشكل التالي: $y = a + b\log x$

6- معادلة الشكل اللوجستي المنطقي والتي يكون لها الشكل التالي: $y = \frac{a_0}{1+ae^{bx}}$

7- أية معادلة أخرى مناسبة لشكل الانتشار.

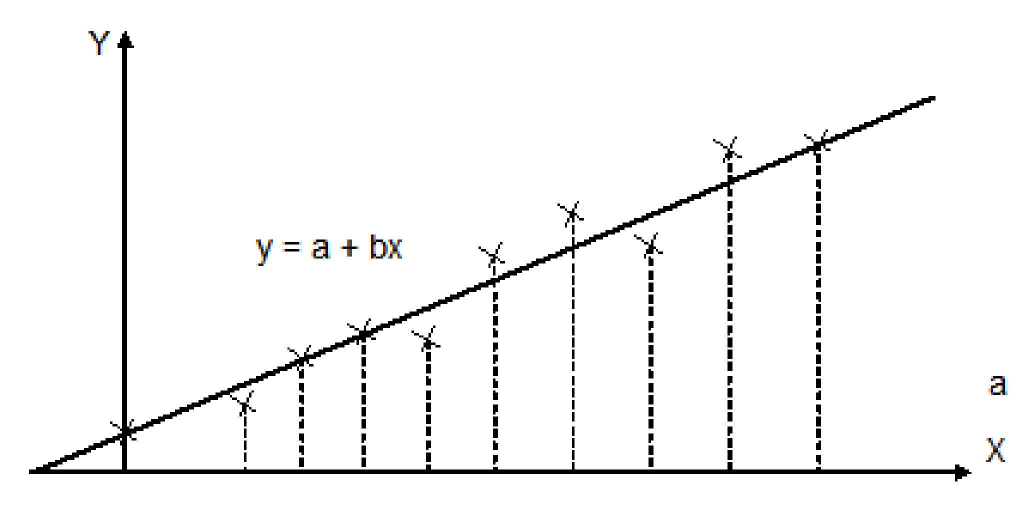
ولكننا هنا سنقتصر على تمثيل العلاقات الارتباطية بواسطة المعادلة الأولى (معادلة الخط المستقيم) ونستعرض ذلك كما يلي:

8-2-1: التمثيل بواسطة معادلة المستقيم:

نقوم بتمثيل العلاقات الارتباطية بين المتحولين y و x بواسطة معادلة المستقيم, إذا كان شكل الانتشار للنقاط الفعلية المتقابلة (x_i , y_i) شبيهاً بخط مستقيم، وعندها نفترض أن معادلة التمثيل (بدلالة القيم المتقابلة) لها الشكل التالي:

$$\tilde{y}_i = a + b\ x_i \qquad (8-8)$$

وإن هذا المستقيم يتوضع على شكل الانتشار ماراً بين النقاط المتقابلة (x_i , y_i) كما يلي:



الشكل (8-9) : معادلة المستقيم على شكل الانتشار

والآن علينا أن نحسب الثابتين a و b (القاطع والميل) بحيث يأخذ ذلك المستقيم أحسن وضع له على شكل الانتشار ويمر بين النقاط بأقل انحرافات ممكنة.

وببساطة نجد أن الوضع المناسب لهذا المستقيم هو الوضع الذي يجعل مجموع مربعات انحرافات القيم الفعلية عن القيم النظرية على المستقيم أقل ما يمكن.

واعتماداً على طريقة المربعات الصغرى (طريقة مشهورة في الإحصاء) نعتبر الثابتين a و b مجهولين ونقوم بحسابهما من المعادلتين العاديتين التاليتين:

$$n\,a + b\sum x_i = \sum y_i$$
$$a\sum x_i + b\sum x_i^2 = \sum x_i\,.\,y_i \qquad (9-8)$$

وهما معادلتان خطيتان تتضمنان المجهولين a و b , ومنهما يمكننا حساب الثابتين a و b فنحصل على القيمتين التقديريتين $a^\backslash$ و $b^\backslash$ لهما, ونحصل على معادلة التمثيل الخطية التالية:

$$\tilde{y}_i = a^\backslash + b^\backslash x_i \qquad (10-8)$$

وهي ترسم خطاً مستقيماً على المستوى XOY ميله يساوي $b^\backslash$ ويقطع المحور OY بالنقطة $(0 , a^\backslash)$

ومن خواص هذا المستقيم (8-10) ما يلي:

1- إن المستقيم (8-10) يمر من النقطة التي إحداثياها مؤلفان من المتوسطين $(\bar{x} , \bar{y})$.

أي أن $\bar{y} = a^\backslash + b^\backslash \bar{x}$

2- إن مجموع القيم النظرية $\tilde{y}_i$ المحسوبة من المستقيم (8-10) يساوي مجموع القيم الفعلية y_i .

أي أن: $\sum_{i=1}^{n} \tilde{y}_i = \sum_{i=1}^{n} y_i$

ويستفاد من هذه الخاصة في التحقق من صحة الحسابات التي قمنا بها.

مثال (8-3): أوجد معادلة المستقيم لتمثيل علاقة دخل الطبيب y مع عدد الزوار x خلال عشرة أيام والتي كانت كما يلي:

رقم اليوم	1	2	3	4	5	6	7	8	9	10
x_i عدد الزوار	5	9	6	8	10	11	12	7	4	3
y_i قيم الدخل (ألف ليرة)	4	5	8	13	15	13	12	9	6	8

الحل : بناءً على شكل الانتشار(8-10), نفترض أن معادلة التمثيل من الشكل الخطي التالي: $\tilde{y}_i = a + b\,x_i$.

ولحساب الثابتين a و b من المعادلتين (8-9)، نعد الجدول التالي لحساب المجاميع $\sum x_i\,y_i$, $\sum x_i$, $\sum y_i$, $\sum x_i^2$ من القيم الفعلية المتوفرة.

رقم اليوم	1	2	3	4	5	6	7	8	9	10	Σ
x_i	5	9	6	8	10	11	12	7	4	3	75
x_i^2	25	81	36	64	100	121	144	49	16	9	645
y_i	4	5	8	13	15	13	12	9	6	8	93
$x_i\, y_i$	20	45	48	104	150	143	144	63	24	24	765

ولحساب الثابتين a و b في معادلة التمثيل الخطي نطبق المعادلتين (8–9) التاليتين :

$$n\,a + b\,\sum x_i = \sum y_i$$
$$a\sum x_i + b\sum x_i^2 = \sum x_i\,.\,y_i$$

وبتعويض قيم المجاميع من الجدول السابق (وn=10) نحصل على المعادلتين التاليتين:

$10\,.\,a + b\,(75) = 93$ (A)

$a\,(75) + b\,(645) = 765$ (B)

وبحساب a من المعادلة (A) وتعويضها في (B) نجد أن: $a = 9.3 - 7.5\,b$

$75\,(9.3 - 7.5\,b) + 645\,b = 765$

$697.5 - 562.5\,b + 645\,b = 765$

$82.5\,b = 67.5$

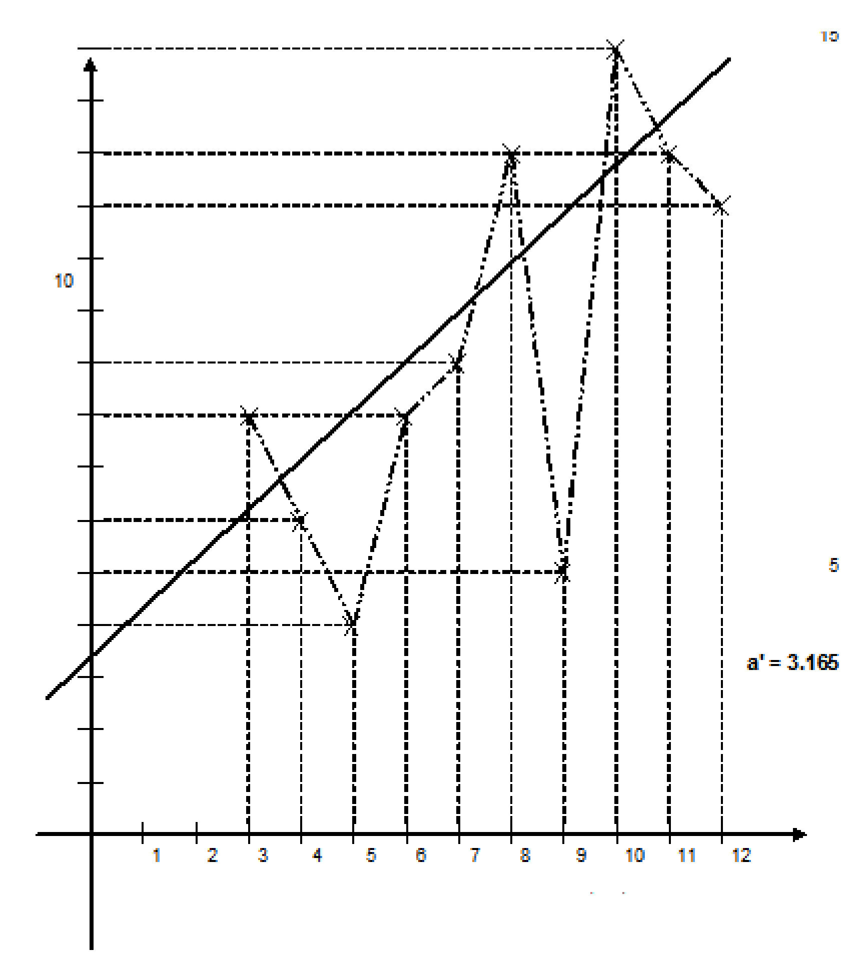
$b^{\backslash} = \frac{67.5}{82.5} = 0.818$

ومنها نحسب تقدير a فنجد أن: $a^{\backslash} = 9.3 - 7.5\,(0.818) = 3.165$

وبذلك نحصل على معادلة المستقيم التي تمثل العلاقة الارتباطية المدروسة والتي تأخذ الشكل التالي:

$\tilde{y}_i = 3.165 + (0.818)\,x_i$ (C)

ويمكننا تمثيل ذلك بيانياً على الشكل التالي:



الشكل (8−10): التمثيل البياني لمعادلة التمثيل

8−2−1: الطريقة المعيارية لحساب ثوابت معادلة التمثيل :

وتعتمد هذه الطريقة على تحويل قيم كل من x و y إلى قيم معيارية وذلك بإجراء التحويلين التاليين:

$$t_i = \frac{x_i - \bar{x}}{\sigma_x}$$ (للمتحول x) (11− 8)

$$Z_i = \frac{y_i - \bar{y}}{\sigma_y}$$ (للمتحول y) (12− 8)

وهذا يعني أنه تم نقل مركز المحاور الإحداثية من النقطة (0 , 0) إلى النقطة $(\bar{x}, \bar{y})$ وتصغير واحدة قياس x بمقدار σ_x وتصغير واحدة قياس y بمقدار σ_y .

وبما أن مستقيم معادلة التمثيل يمر من النقطة $(\bar{x}, \bar{y})$ والتي أصبحت مركزاً للإحداثيات فهذا يعني أن الحد الثابت a في معادلة المستقيم أصبح يساوي صفراً (a = 0).

أي أن معادلة التمثيل الجديدة لم تعد تتضمن حداً ثابتاً مثل a , لذلك فإنه يمكننا افتراض أن معادلة التمثيل المعيارية بين المتحولين المعياريين الجديدين t_i و Z_i تأخذ الشكل التالي:

$$\tilde{Z}_i = \beta\, t_i$$ (13− 8)

ولحساب قيمة الثابت β نطبق المعادلة التالية:

$$\beta = \frac{\sum_{i=1}^{n}(Z_i * t_i)}{\sum t_i^2} \qquad (14-8)$$

فنحصل على قيمة تقديرية لها هي $\beta^\backslash$ وبالتالي نحصل على معادلة التمثيل المعيارية التالية:

$$\tilde{Z}_i = \beta^\backslash * t_i \qquad (15-8)$$

ويمكن البرهان على أن قيمة $\beta^\backslash$ تساوي معامل الارتباط r ، أي أن:

$$\beta^\backslash = r \qquad (16-8)$$

وإن قيمة $\beta^\backslash$ تظهر في جميع النتائج التي يقوم بها الحاسب عند حساب معادلة التمثيل.

مثال (*8-4*) : أوجد معادلة التمثيل الخطي المعيارية لعلاقة عدد الزوار x لطبيب الأسنان بقيمة دخله اليومي y ، وذلك من خلال المعلومات المأخوذة من السبعة أيام التالية:

اليوم i	1	2	3	4	5	6	7	Σ	المتوسط	التباين
عدد الزوار x_i	2	4	6	8	10	12	14	56	8	16
قيمة الدخل y_i (ألف ليرة)	5	7	9	10	11	13	15	70	10	10

الحل : من خلال المعلومات المبينة أعلاه نجد أن متوسطي x و y يساويان:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{56}{7} = 8$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{70}{7} = 10$$

كما نجد أن تباينيهما σ_x^2 و σ_y^2 يساويان:

$$\sigma_x^2 = \frac{\sum(x_i - 8)^2}{7} = \frac{36+16+4+0+4+16+36}{7} = 16$$

$$\sigma_y^2 = \frac{\sum(y_i - 10)^2}{7} = \frac{25+9+1+0+1+9+25}{7} = 10$$

أي أن : $\sigma_x = \sqrt{\sigma_x^2} = \sqrt{16} = 4\sigma_y = \sqrt{\sigma_y^2} = \sqrt{10} = 3.16$

لإيجاد معادلة التمثيل المعيارية لا بد من أن نقوم بحساب المتحولين المعياريين:

$$Z_i = \frac{y_i - \bar{y}}{\sigma_y} \qquad , \qquad t_i = \frac{x_i - \bar{x}}{\sigma_x}$$

ثم حساب الجداءات $Ƶ_i * t_i$ وحساب قيم t_i^2 . ولتسهيل هذه العمليات نعد الجدول التالي:

I	1	2	3	4	5	6	7	Σ
$t_i = \frac{x_i - \bar{x}}{\sigma_x}$	−1.5	−1	−0.5	0	0.5	1	1.5	0
$Ƶ_i = \frac{y_i - \bar{y}}{\sigma_y}$	−1.58	−0.95	−0.316	0	0.316	0.95	1.58	0
* $Ƶ_i t_i$	2.37	0.95	0.158	0	0.158	0.95	2.37	6.956
t_i^2	2.25	1	0.25	0	0.25	1	2.25	7

وبالتعويض في المعادلة (8−14) نحصل على قيمة تقديرية لـ β هي:

$$\beta^\backslash = \frac{\sum Ƶ_i\, t_i}{\sum t_i^2} = \frac{6.956}{7} = 0.9932$$

وبذلك نحصل على معادلة التمثيل المعيارية التالية:

$$\tilde{Ƶ}_i = 0.9932\, t_i$$

ومنها يمكننا أن نستنتج مباشرة قيمة معامل الارتباط بين المتحولين Ƶ و t أو بين المتحولين y و x وهي:

$$r_{xy} = \beta^\backslash = 0.9932$$

وهذا يدل على وجود ارتباط طردي متين.

وأخيراً يمكننا العودة إلى المتحولين الأساسيين y و x والحصول على المعادلة الأساسية لتمثيل العلاقة بين المتحولين y و x من المعادلة المعيارية وذلك باستبدال كل من $Ƶ_i$ و t_i بما تساويه , فنجد من المعادلة المعيارية أن:

$$\frac{\tilde{y}_i - 10}{3.16} = 0.9932\left(\frac{x_i - 8}{4}\right)$$

$$\tilde{y}_i = 10 + \frac{3.16\,(0.9932)}{4}\, x_i - \frac{3.16\,(0.9932)\,.\,8}{4}$$

$$\tilde{y}_i = 3.714 + 0.786\, x_i.$$

8−3: دراسة جودة التمثيل للمعادلة المحسوبة :

لقد لاحظنا من خلال شكل الانتشار (8−10) أن الخط المستقيم الذي يعبر عن معادلة التمثيل يمر بين النقاط الفعلية، وهذا يعني أن قيم Y الفعلية y_i تختلف عن القيم النظرية المحسوبة من المعادلة $\tilde{y}_i$، وكلما كان هذا الاختلاف صغيراً كان تمثيل المعادلة المحسوبة جيداً وقوياً. ولحساب مقدار هذا الاختلاف نقوم بحساب ما يسمى بتباين التمثيل (أو بالتباين غير المفسر) من العلاقة:

$$S^2 = \frac{1}{n-2}\sum (y_i - \tilde{y}_i)^2 \qquad (17-8)$$

ثم نقوم بحساب التباين الكلي لـ y *من العلاقة:*

$$\sigma_y^2 = \frac{1}{n}\sum(y_i - \bar{y}_i)^2 \qquad (18-8)$$

ثم نحسب معامل التحديد لقياس جودة التمثيل من العلاقة:

$$R^2 = 1 - \frac{S^2}{\sigma_y^2} \qquad (19-8)$$

وكلما كانت قيمة R^2 قريبة من الواحد كان التمثيل جيداً، وعندها نقول أن تغيرات X تفسر لنا تغييرات Y بنسبة $100 * R^2$ %.

ملاحظة: يمكن البرهان على أنه إذا كانت معادلة التمثيل خطية فإن قيمة R^2 تساوي مربع معامل الارتباط للمتحولين Y *و* X ، *أي أن :*

$$R^2 = r^2 \qquad (20-8)$$

كما يوجد قيمة مصححة لـ R^2 تستخدم للارتباط المتعدد الخطي وهي تساوي:

$$Adjusted\ R^2 = 1 - \frac{(n-1)}{n-(k+1)}(1-R^2) \qquad (21-8)$$

حيث أن k عدد المتحولات المستقلة في معادلة الارتباط المتعدد الخطي.

مثال(8−5): لدراسة جودة تمثيل المعادلة التي حصلنا عليها في المثال (8−4), نقوم بإيجاد القيم النظرية لـ $\tilde{y}_i$ من معادلة التمثيل المحسوبة، ثم نقوم بحساب مجموع مربعات انحرافات القيم النظرية $\tilde{y}_i$عن القيم الفعلية المقابلة y_i لذلك نعد الجدول المساعد التالي:

/	1	2	3	4	5	6	7	$\sum$
القيم الفعلية y_i	5	7	9	10	11	13	15	70
القيم النظرية المحسوبة من المعادلة $\tilde{y}_i = 0.786.x_i + 3.7161$	5.29	6.86	8.43	10	11.57	13.14	14.71	70
$y_i - \tilde{y}_i$	−0.29	0.14	0.57	0	0.57	−0.14	0.29	0
$(y_i - \tilde{y}_i)^2$	0.0841	0.0196	0.325	0	0.325	0.0196	0.0841	0.8574

ومن نتائج حسابات هذا الجدول نجد أن تباين التمثيل (التباين غير المفسر) يساوي :

$$S^2 = \frac{\sum(y_i - \tilde{y}_i)^2}{n-2} = \frac{0.8574}{5} = 0.17148$$

ولقد وجدنا سابقاً أن التباين الكلي لـ Y يساوي : $\sigma_y^2 = 10$.

وبذلك نجد أن قيمة معامل التحديد R^2 تساوي:

$$R^2 = 1 - \frac{S^2}{\sigma_y^2} = 1 - \frac{0.17148}{10} = 0.9829$$

وهي قيمة قريبة جداً من الواحد، وهي تدل على أن جودة تمثيل المعادلة $\tilde{y}_i = \mathbf{0.786\,x_i + 3.714}$ للعلاقة بين Y و X هو تمثيل جيد جداً. ومن قيمة R^2 نجد أن تغيرات X تفسر 98.29% من تغيرات Y , وإن الباقي يعود إلى تغيرات المتحولات الأخرى غير الداخله في معادلة التمثيل.

4-8 : التنبؤ بواسطة معادلة التمثيل :

تستخدم معادلة التمثيل للتنبؤ بقيم Y المقابلة لقيم معلومة لـ X , فإذا أخذنا أي قيمة لـ X غير موجودة في الجدول السابق مثل x=16 فإنه يمكننا التنبؤ بقيمة Y المقابلة لها من المعادلة السابقة كما يلي:

$$\tilde{y}_{(16)} = 0.876(16) + 3.714 = 16.29$$

وهو قيمة تقديرية للقيمة الحقيقية Y المقابلة لـ x=16, ولكنها تختلف عنها وتتضمن خطأ بالزيادة أو النقصان. ويقدر مقدار هذا الخطأ بواسطة حساب الانحراف المعياري لخطأ التمثيل من(8−17) وهو عبارة عن المقدار s , الذي يسمى بالخطأ المعياري للتمثيل، والذي يحسب من العلاقة التالية:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum(y_i - \tilde{y}_i)^2}{n-2}}$$

وبناء على ذلك يمكننا حساب مجال الثقة الثاني للقيمة الحقيقية y_{16} بواسطة العلاقة التقريبية التالية:

$$p\left[\tilde{y}_{(16)} - 2.s \le y_{16} \le \tilde{y}_{(16)} + 2.s\right] = 0.095$$

وكتطبيق على ذلك نقوم بحساب مجال الثقة للقيمة الحقيقية y_{16} فنجد أن الخطأ المعياري يساوي:

$$s = \sqrt{0.1748} = 0.4141$$

وإن مجال الثقة الثاني لها يساوي:

$$\boldsymbol{p[16.29 - 2(0.4141) \le y_{16} \le 16.29 + 2(0.4141)] = p[15.46 \le y_{16} \le 17.12] = 0.95}$$

أي أن القيمة الحقيقية لـ y_{16} تكون محصورة بين القيمتين 15.46 و 17.2 باحتمال ثقة قدره 0.95 .

ملاحظة : يمكن تعميم فكرة الارتباط البسيط إلى الارتباط المتعدد وإدخال عدد متحولات مستقلة في معادلة التمثيل ، وعندها نكتب ذلك كما يلي:

$$\tilde{y}_i = a_0 + a_1 x_{1i} + a_2 x_{2i} + a_3 x_{3i}$$

وإن حساب الثوابت: a_0 و a_1 و a_2 و a_3 تحتاج إلى معادلات خاصة مشابهة للمعادلتين (8− 9) السابقتين.

8-5: السلاسل الزمنية:

السلسلة الزمنية هي عبارة عن متوالية من القيم العددية لأحد المتحولات y مقرونة بالأزمنة t المقابلة لها , ونكتبها على الشكل التالي :

$$\begin{matrix} t: & t_1 & t_2 & t_3 & \ldots\ldots & t_i & \ldots\ldots & t_n \\ y: & y_1 & y_2 & y_3 & \ldots\ldots & y_i & \ldots\ldots & y_n \end{matrix} \qquad (22-8)$$

وإن معالجة هذه السلاسل لا تختلف كثيراً عن معالجة الارتباط والانحدار بين المتحولين الكميين y و x . إلا من ناحية واحدة وهي أن الزمن t لا يعتبر عاملاً مسبباً للظواهر بل مرافقاً لها .

ويمكن أن تمثل العلاقة بين المتحول y والزمن t بواسطة أي من المعادلات أو النماذج, التي ذكرناها سابقاً, ثم نقوم بدراسة جودة التمثيل والتنبؤ بقيم y المستقبلية كما فعلنا في حالة المتحولين x و y .

8-5-1: المؤشرات الإحصائية للسلاسل الزمنية:

1- التغير السنوي : وهو يعبر عن الزيادة أو النقصان بين أي قيمتين متتاليتين أو يساوي الفرق بين كل قيمة y_i عن سابقتها y_{i-1} ويحسب من العلاقة:

$$\Delta_i = y_i - y_{i-1} \qquad (23-8)$$

2- التغير الكلي: وهو الفرق بين القيمة الأخيرة y_n والقيمة الأولى y_1 ويحسب من العلاقة:

$$D = y_n - y_1 \qquad (24-8)$$

3- الأرقام القياسية الثابتة: وهي عبارة نسبة كل قيمة y_i على القيمة الأولى y_1 مضروبة بـ 100 بالمئة , وتكتب على الشكل التالي:

$$\frac{y_1}{y_1}100\ ,\ \frac{y_2}{y_1}100\ ,\ \frac{y_3}{y_1}100\ ,\ \frac{y_4}{y_1}100\ \ldots\ldots\frac{y_n}{y_1}100 \qquad (25-8)$$

4- الأرقام القياسية المتحركة: وهي عبارة عن نسبة كل قيمة y_i على سابقتها y_{i-1} مضروبة بـ100 بالمئة, وتكتب على الشكل التالي:

$$\frac{y_2}{y_1}100\,,\frac{y_3}{y_2}100\,,\frac{y_4}{y_3}100\,,\ldots\ldots\ \frac{y_n}{y_{n-1}}100 \qquad (26-8)$$

وإذا رمزنا لهذه الأرقام بـ I_1 ، I_2 ، I_3 .... I_{n-1}

فإن متوسط هذه الأرقام القياسية المتسلسلة يحسب بواسطة المتوسط الهندسي لها، أي أن :

$$\bar{I} = \sqrt[n-1]{I_2 \,.\, I_3 \,.\, I_4 \ldots .\, I_{n-1}} = \sqrt[n-1]{\frac{y_n}{y_1}}.\,100\% \qquad (27-8)$$

5- معدل النمو السنوي: وهو عبارة عن نسبة التغيير السنوي على القيمة السابقة y_{i-1} مضروباً بـ 100 بالمئة , ويحسب من العلاقة:

$$T_i = \frac{y_i - y_{i-1}}{y_{i-1}} 100 = \frac{\Delta_i}{y_{i-1}} 100 \qquad (28-8)$$

وهنا يمكننا استنباط علاقة هامة تربط معدل النمو السنوي بالرقم القياسي المتسلسل، حيث نجد من العلاقة (8-28) أن:

$$T_i = \frac{y_i}{y_{i-1}} 100 - 100 = I_i - 100 \qquad (29-8)$$

6- إذا أردنا حساب متوسط معدل النمو السنوي خلال كامل الفترة (t_1 , t_n) فإننا سنميز بين حالتين هما:

a. إذا كان تطور الظاهرة يتم على شكل خط مستقيم، فإننا نحسب متوسط معدل النمو السنوي من العلاقة:

$$\bar{T} = \frac{(y_n - y_1)}{(n-1)y_1} 100 \qquad (30-8)$$

b. إذا كان تطور الظاهرة يتم بشكل اسي، فإننا نحسب متوسط معدل النمو السنوي من العلاقة التي تعطينا الفائدة المركبة بمعدل $\frac{\bar{T}}{100}$ ولمدة (n-1) سنة ويكون لدينا :

$$y_n = y_1 \left(1 + \frac{\bar{T}}{100}\right)^{n-1}$$

وبذلك نجد أن :

$$1 + \frac{\bar{T}}{100} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

$$\frac{\bar{T}}{100} = \sqrt[n-1]{\frac{y_n}{y_1}} - 1 \qquad (31-8)$$

$$\bar{T} = \sqrt[n-1]{\frac{y_n}{y_1}} \times 100 - 100$$

$$\bar{T} = \bar{I} - 100 \qquad (32-8)$$

ملاحظة: إن تحليل السلاسل الزمنية يتناول جميع مركباتها(الاتجاه العام والتغيرات الموسمية والتقلبات الدورية والأخطاء العشوائية)، وهي أمور طويلة ومعقدة وتحتاج إلى دراسات خاصة بذلك.

مثال (8–6) : لنفترض أن عدد سكان إحدى المدن تطور خلال خمس سنوات كما يلي :

العام	2000	2001	2002	2003	2004	2005
عدد السكان في منتصف العام	1200	1350	1470	1630	1710	1800
الزمن t	1	2	3	4	5	6

والمطلوب : حساب المؤشرات الإحصائية لهذه السلسلة الزمنية :

الحل: نكتب السلسلة الزمنية بشكل مختصر ثم نقوم بحساب مؤشراتها ضمن جدول كما يلي :

الزمن t_i	**1**	**2**	**3**	**4**	**5**	**6**
عدد السكان y_i	**1200**	**1350**	**1470**	**1630**	**1710**	**1800**
$\Delta_i = y_i - y_{i-1}$	–	**150**	**120**	**160**	**80**	**90**
الأرقام القياسية الثابتة $\bar{I}_i = \frac{y_i}{y_1}100$	**100**	**112.5**	**122.5**	**135.8**	**142.5**	**150.0**
الأرقام القياسية المتحركة $\bar{I}_i = \frac{y_i}{y_{i-1}}100$	–	**112.5**	**108.9**	**110.9**	**104.9**	**105.3**
معدلات النمو السنوية $T_i = I_i - 100$	–	**12.5**	**8.9**	**10.9**	**4.4**	**5.3**

وبذلك نجد أن متوسط الأرقام القياسية المتحركة يساوي :

$$\bar{I} = \sqrt[5]{I_1 \times I_2 \times I_3 \times I_4 \times I_4} = 108.46$$

وأخيرا نجد أن متوسط معدل النمو السنوي يساوي :

$$\bar{T} = \frac{(y_n - y_1)}{(n-1)y_1}100 = \frac{(1800 - 1200)}{5(1200)}100 = 10\%$$ (للتطور الخطي)

$$\bar{T} = \sqrt[n-1]{\frac{y_n}{y_1}} - 100 - 100 = 8.45\%$$ (للتطور الأسي)

وكان يمكن حساب المعدل الأخير من العلاقة :

$$\bar{T} = \bar{I} - 100 = 108.45 - 100 = 8.46\%$$

8—6 : اختبار معنوية معامل الارتباط الخطي r_{xy}:

لقد عرفنا في بداية هذا الفصل أن معامل الارتباط الخطي بين متحولين x , y بالعلاقة التالية:

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n.\,\sigma_x\,\sigma_y} \qquad (33-8)$$

ولم نشر وقتها إلى أن هذا المعامل يختلف من عينة لأخرى، لذلك يمكن اعتباره متحولاً عشوائيا ويخضع لتوزيع احتمالي معين، وإن متوسطه أو مركزه يساوي قيمة ما ρ وتباينه يساوي:

$$\sigma_r^2 = \frac{1 - r^2}{n - 2} \qquad (34-8)$$

وانحرافه المعياري يساوي:

$$\sigma_r = \sqrt{\frac{1 - r^2}{n - 2}} \qquad (35-8)$$

وعند اختبار معنوية معامل الارتباط r_{xy} يجب علينا أن نحدد التوزيع الاحتمالي الذي يخضع له، كما يقتضي الأمر تعريف مؤشر الاختبار اللازم لذلك، والتوزيع الاحتمالي الخاضع له. وهنا نميز بين حالتين أساسيتين هما:

1- الحالة التي يفترض أن يكون فيها $\rho = 0$:

وفي هذه الحالة يكون التوزيع الاحتمالي لـ r متناظراً حول القيمة $\rho = 0$. ويمكن اعتباره توزيعاً مقارباً للتوزيع الطبيعي العام. وبذلك يمكننا تعريف مؤشر الاختبار بواسطة المتحول المعياري التالي:

$$t = \frac{r - \rho}{\sigma_r} = \frac{r - \rho}{\sqrt{\frac{1 - r^2}{n - 2}}} \qquad (36-8)$$

وهو متحول يخضع لتوزيع ستودينت t بـ (n−2) درجة حرية. وعند إجراء الاختبار نقوم بوضع الفرضيتين العدمية والبديلة كما يلي:

$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$

ثم نحسب قيمة مؤشر الاختبار t من العلاقة (8−36) ونقارنها مع القيمة الجدولية t_0 المقابلة لـ (n−2) درجة حرية ولنصف مستوى الدلالة $\alpha/2$. فإذا كانت القيمة المطلقة لـ t المحسوبة أصغر من t_0 الجدولية, أي إذا كانت : $|t| < t_0$, فإننا نقبل الفرضية العدمية H_0, التي تقول بعدم وجود ارتباط بين المتحولين x,y ، أما إذا كانت $|t| > t_0$ فإننا نرفض H_0 ونقبل الفرضية البديلة, التي تقول أن قيمة معامل الارتباط r معنوية ونعترف بوجود ارتباط بين المتحولين x , y .

مثـال(8−7) : لنفتـرض أن عينـة مؤلفـة مـن $n = 10$ أزواج مـن الـذكور والإنـاث, أظهـرت أن معامـل الارتباط بين أطوالهم كان $r = 0.70$, والمطلوب اختبار معنوية هذا المعامل بمستوى دلالة $\alpha = 0.05$

<u>الحل:</u> نضع الفرضيتين كما يلي: $H_0 : \rho = 0$ ، $H_1 : \rho \neq 0$ ويكون الاختبار ثنائي الجانب.

ثم نحسب قيمة مؤشر الاختبار t من العلاقة:

$$t = \frac{r - \rho}{\sqrt{\frac{1 - r^2}{n - 2}}} = \frac{0.70 - 0}{\sqrt{\frac{1 - (0.70)^2}{10 - 2}}} = 2.72$$

ومن جداول توزيع ستودينت نجد أن قيمة t المقابلة لـ $n-2 = 10-2 = 8$ درجة حرية ولنصف مستوى الدلالة $\frac{\alpha}{2} = 0.025$ ، تساوي $t_{0.025} = 2.306$. وبمقارنة القيمة المحسوبة لـ t مع القيمة الحرجة t_0 نجد أن $|t| > t_0$ ، لذلك نرفض الفرضية H_0، ونعتبر أن قيمة معامل الارتباط r هي قيمة معنوية أو حقيقية وإن الارتباط بين أطوال الذكور والإناث موجود وله قيمة معنوية.

2− الحالة التي يفترض أن يكون فيها $\rho \neq 0$:

وفي هذه الحالة يكون التوزيع الاحتمالي لـ r_{xy} ملتوياً نحو اليمين أو اليسار ولا يجوز في هذه الحالة استخدام توزيع ستودينت t. وهنا نضع فرضية العدم كما يلي: $H_0 : \rho = \rho_0$، حيث هي القيمة المتوقعة لمعامل الارتباط في المجتمع، وتحدد من قبل الباحث مسبقاً.

وللتخلص من هذه المشكلة نقوم بتحويل r إلى متحول طبيعي بشكل تقريبي بواسطة التحويل التالي:

$$Z_r = \frac{1}{2} \ln \frac{1 + r}{1 - r} \qquad (37 - 8)$$

وتم البرهان على أن مركزه أو متوسطه $\overline{Z}_r$ يساوي:

$$\overline{Z}_r = \frac{1}{2} \ln \frac{1 + \rho}{1 - \rho} \qquad (38 - 8)$$

وعلى أن تباينه يساوي:

$$\sigma^2_{z_r} = \frac{1}{n - 3} \qquad (39 - 8)$$

وانحرافه المعياري يساوي:

$$\sigma_{z_r} = \sqrt{\frac{1}{n - 3}}$$

وهكذا نجد أنه يمكننا تعريف مؤشر اختبار خاضع للتوزيع الطبيعي المعياري بواسطة العلاقة:

$$Z = \frac{Z_r - \overline{Z}_r}{\sigma_{z_r}} = \frac{Z_r - \overline{Z}_r}{\sqrt{\frac{1}{n - 3}}} \qquad (40 - 8)$$

وبعد حساب قيمة Z نقارنها مع قيمة Z_0 الحرجة المقابلة لنصف مستوى الدلالة $\frac{\alpha}{2}$، ونتخذ القرار كما يلي:

إذا كانت $|Z| < Z_0$ نقبل فرضية العدم $H_0 : \rho = \rho_0$.

أما إذا كانت $|Z| \geq Z_0$ نرفض فرضية العدم $H_0 : \rho = \rho_0$.

مثال(8-8) : لنفترض أن عينة بحجم n=12 أسرة أظهرت أن قيمة معامل الارتباط بين أعمار الزوجين كانت r=0.70 ، اختبر معنوية هذا المعامل ضد الفرضية $\rho = 0.50$ وبمستوى دلالة $\alpha = 0.05$.

الحل: في هذه الحالة نضع الفرضيتين العدمية والبديلة كما يلي:

$$H_0 : \rho = 0.50$$
$$H_1 : \rho \neq 0.50$$

ثم نقوم بحساب Z_r , $\overline{Z}_r$ من العلاقتين السابقتين فنجد أن:

$$Z_r = \frac{1}{2} \ell_n \frac{1 + 0.70}{1 - 0.70} = 0.867$$

$$\overline{Z}_r = \frac{1}{2} \ell_n \frac{1 + 0.5}{1 - 0.5} = 0.549$$

ثم نحسب قيمة مؤشر الاختبار Z من العلاقة:

$$Z = \frac{0.867 - 0.549}{\sqrt{\frac{1}{12 - 3}}} = 0.954$$

ثم نبحث في جدول التوزيع الطبيعي عن قيمة Z_0 الحرجة المقابلة لنصف مستوى الدلالة $\frac{\alpha}{2} = 0.025$ ، فنجد أنها تساوي $Z_0 = 1.96$. ثم نقارن قيمة Z المحسوبة مع قيمة Z_0 الحرجة، فنجد أن: $|Z| < Z_0$. لذلك نقبل فرضية العدم والتي تقول أن معامل الارتباط يساوي $\rho = 0.5$.

7-8 : اختبار معنوية معامل الارتباط الرتبي r_s:

لقد عرفنا سابقآ أن معامل الارتباط الرتبي بالعلاقة:

$$r_s = 1 - \frac{6 \sum (k_i - p_i)^2}{n(n^2 - 1)}$$

ولإجراء اختبار حول معنوية هذا المعامل تستخدم جداول خاصة بذلك وخاصة عندما تكون : $n < 20$ ولن نتعرض لها في هذا الفصل . ولكن عندما يكون حجم العينة $n \geq 20$ ، فإنه يمكننا استخدام مؤشر الاختبار التالي:

$$Z = \frac{r_s - \rho}{\sigma_{r_s}} = \frac{r_s - \rho}{\sqrt{\frac{1}{n-1}}} \qquad (41-8)$$

وذلك لأن :
$$\sigma_r^2 = \frac{1}{n-1}$$

مثال(8-9) : في دراسة لـ 26 أسرة وجدنا أن معامل الارتباط الرتبي يبين مستوى تعليم الزوج والزوجة كان $r_s = 0.70$. والمطلوب: اختبار معنوية هذا المعامل بمستوى دلالة $\alpha = 0.05$.

<u>الحل:</u> نضع الفرضيتين كما يلي:

$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$

ثم نحسب قيمة مؤشر الاختبار Z:

$$Z = \frac{r_s - \rho}{\sqrt{\frac{1}{n-1}}} = \frac{0.70 - 0}{\sqrt{\frac{1}{25}}} = 3.5$$

ومن جدول التوزيع الطبيعي نجد أن قيمة Z_0 االحرجة المقابلة لنصف مستوى الدلالة $\frac{\alpha}{2} = 0.025$ ، تساوي $Z_0 = Z_{1-\frac{\alpha}{2}} = 1.96$. وبمقارنة هاتين القمتين نجد أن $|Z| > Z_0$ ، لذلك نرفض الفرضية H_0 ونقبل الفرضية H_1 والتي تقول أن قيمة معامل الارتباط الرتبي $r = 0.70$ هي قيمة معنوية وتعبر عن وجود ارتباط بين مستوى التعليم للزوج والزوجة.

تمرينات

1- لنفترض أنه لدينا المعلومات التالية عن عدد السكان من القطر وقيمة الناتج الإجمالي (مليار ل.س) خلال الفترة 1980-1986 :

العام	1980	1981	1982	1983	1984	1985	1986
عدد السكان x مليون نسمة	8	9	9.3	9.6	10	10.3	10.6
الناتج الاجمالي y مليار ل.س	51	56	58	59	57	58.5	58.7

المصدر : المجموعة الإحصائية 1988 ص 5-502

والمطلوب : 1- رسم شكل الانتشار لـ y و x واقتراح نوع معادلة التمثيل .

2- حساب معامل الارتباط بين y و x .

3- ايجاد معادلة التمثيل الخطية بين y و x .

4- دراسة جودة التمثيل للمعادلة السابقة .

5- التنبؤ بقيمة الناتج عندما يصبح عدد السكان 12 مليونا

2- لنفترض أنه لدينا البيانات التالية عن تطور سكان مدينة ما :

العام t	1995	1996	1997	1998	1999	2000	2001	2002
عدد السكان بالآلاف y_t	380	410	450	490	520	550	600	640

المطلوب : 1- حساب التغيرات السنوية .

2- حساب التغير الكلي خلال الفترة .

3- حساب الأرقام القياسية الثابتة .

4- حساب الأرقام القياسية المتسلسل .

5- حساب متوسط معدل النمو السنوي .

6- حساب معامل الارتباط .

7- رسم شكل الانتشار واقتراح معادلة لتمثيل تطور السكان مع الزمن .

8- ايجاد معادلة التمثيل الخطية ودراسة جودة تمثيلها .

9- التنبؤ بعدد السكان في عام 2005 وفي عام 2010 .

10- ايجاد مجال الثقة للقيم المتنبأ بها باحتمال ثقة 0.95 .

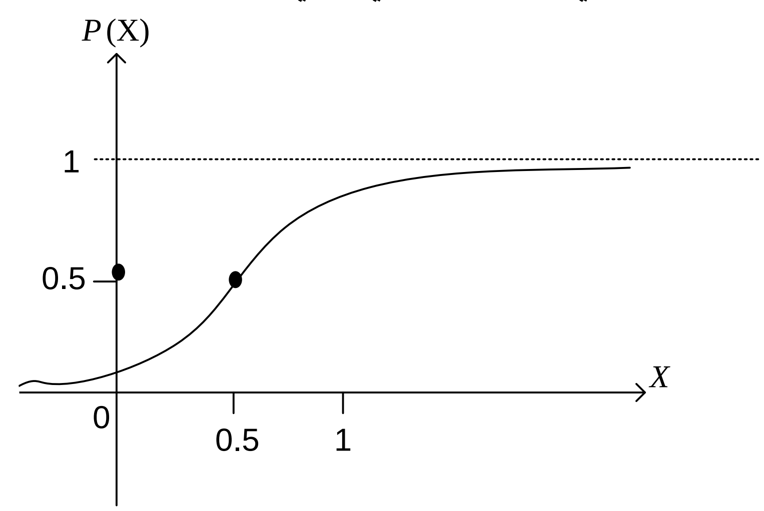
الفصل التاسع
التحليل اللوجستي

9-1- تمهيد:

يهدف التحليل اللوجستي إلى تصنيف عناصر المجتمع المدروس إلى مجموعتين أو أكثر، وذلك باستخدام التوزيع الاحتمالي اللوجستي المعرف بالعلاقة التالية:

$$P(X) = \frac{1}{1 + \bar{e}^{(\beta_0 + \beta_X)}} \quad : \quad -\infty < X < \infty \qquad (1-9)$$

حيث X هو شعاع المتحولات المؤثرة في عمليات التصنيف و$P(x)$ هو الاحتمال المقابل له ويأخذ قيمه في المجال [0,1]، وهو يرسم في المستوى المنحني التالي:



الشكل (9-1): منحني التوزيع اللوجستي

وهناك نوعان للتحليل اللوجستي هما: الثنائي والمتعدد :

1- **التحليل اللوجستي الثنائي:** وفيه يشترط أن يكون التابع Y نوعياً ويأخذ حالتين متنافيتين (نجاح أو فشل، ربح أو خسارة، مدخن أو غير مدخن، حامل للمرض أو غير حامل له، محقق لشرط ما أو غير محقق له، ...الخ)، وأن يأخذ مقابل الحالة المرغوبة الأولى القيمة العددية (1) واحد، وأن يأخذ مقابل الحالة الثانية القيمة العددية (0) صفر. أما المتحولات X المؤثرة في التصنيف فيمكن أن تكون كمية أو نوعية أو مختلطة، وتأخذ قيمها ضمن مجالات أو فئات محددة، ولا يشترط عليها أن تحقق أية شروط مسبقة.

2- **التحليل اللوجستي المتعدد:** وفيه يشترط أن يكون التابع Y نوعياً ويأخذ عدة حالات متنافية (مستوى التعليم، حالة العمل، الحالة الاجتماعية، ...الخ) وأن يأخذ مقابل إحدى الفئات القيمة (1) ومقابل الفئات المتبقية القيمة (0).

مثال (9−1): لنفترض أن دراسة بواسطة الاستبيان شملت (20) طالباً، لمعرفة علاقة متوسط عدد ساعات الدراسة اليومية X مع نتيجة اجتيازهم للامتحان Y، الذي يأخذ القيمة (1) في حالة النجاح والقيمة (0) في حالة الرسوب. وكانت نتائج الاستبيان كما في الجدول التالي:

جدول (9−1): نتائج الاستبيان لعلاقة عدد الساعات بنتيجة الامتحان [Wikipedia.org]:

i رقم الطالب	1	2	3	4	5	6	7	8	9	10
X: عدد الساعات	0.50	0.75	1.00	1.25	1.50	1.75	1.75	2.00	2.25	2.50
Y: نتيجة الامتحان	0	0	0	0	0	0	1	0	1	0

يتبع

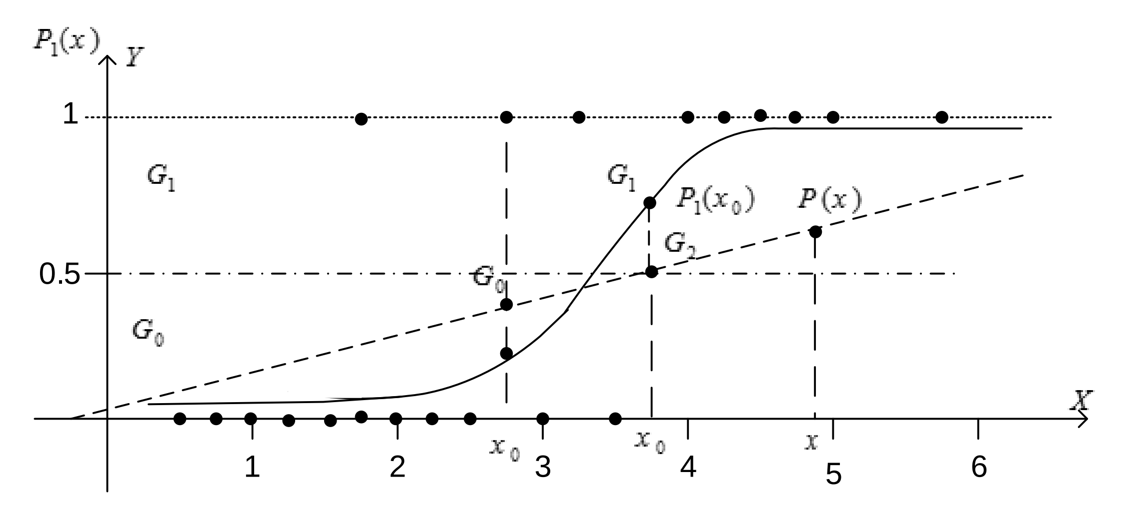
i رقم الطالب	11	12	13	14	15	16	17	18	19	20
X: عدد الساعات	2.75	3.00	3.25	3.50	4.00	4.25	4.50	4.75	5.00	5.50
Y: نتيجة الامتحان	1	0	1	0	1	1	1	1	1	1

ونريد الآن معرفة مدى تأثير عدد ساعات الدراسة على احتمال النجاح.

ومن الجدول السابق نلاحظ أن عدد الناجحين $m = 10$، أي أن المعدل العام للنجاح يساوي: $P = \frac{10}{20} = 0.50$، وبالمقابل نجد أن المعدل العام للرسوب $q = 0.50$.

وإذا أردنا رسم شكل الانتشار لهذه البيانات، فإننا نلاحظ أن المتحول المستقل X هو متحول مستمر ويأخذ قيمه في المجال [6 , 0].

أما المتحول التابع Y فهو متحول منقطع وثنائي القيمة، فهو يأخذ القيمة $(Y = 1)$ في حالة النجاح، ويأخذ القيمة (0) في حالة الرسوب. وإذا قمنا برسم النقاط (x_1, y_1) على المستوى نجد أن قيم Y الصفرية تتوضع على المحور $0X$ وتميل نحو الجانب الأيسر، أما قيم Y المساوية للواحد فتتوضع على المستقيم$(Y = 1)$ وتميل نحو الجانب الأيمن، ويوجد بعض النقاط المتقابلة في المنطقة الوسطى كما هو مبين على الشكل التالي:



الشكل (9-2): شكل الانتشار للبيانات الثنائية والمنحني اللوجستي

والسؤال الآن كيف سنتعامل مع هذا الشكل العجيب؟

وللإجابة على هذا السؤال سندرس أولاً بعض خصائص الظواهر الثنائية من خلال بيانات المثال (9-1) السابق.

9-2: خواص الظواهر الثنائية:

إن الظواهر الثنائية هي عبارة عن توابع ثنائية تأخذ حالتين A و$\bar{A}$ فقط [كأن تأخذ: موافق أو غير موافق، نعم أو لا، نجاح أو فشل، ربح أو خسارة، قبول أو رفض، ...الخ] . ولقد أصطلح على إعطاء التابع الثنائي Y قيمة الواحد (1) عندما تتحقق الحالة المرغوبة A، وقيمة الصفر (0) عندما تحقق الحالة غير المرغوبة $\bar{A}$ (عدم تحقق A). ولنفترض أنه عند إجراء (n=100) تجربة على إحدى الظواهر الثنائية كانت نتائج تلك التجارب التي تخضع لتوزيع (برنويللي) كما في بالجدول التالي:

جدول (9-2): مخطط جدولي لتوزيع (برنولي) لـ 100 تجربة على Y.

الحالة	$A = G_1$	$\bar{A} = G_0$	المجموع
قيمة التابع Y	1	0	----
احتمال التحقيق	p	q	$p + q = 1$
عدد التكرارات المطلقة	n_1	n_0	$n_1 + n_0 = n$
توزع عدد التجارب	60	40	$100 = n$

ومن هذه التجربة سنحصل على عينة من قيم Y بحجم n ، وإن العناصر المقابلة لها تتوزع حسب الجدول (9-2) على مجموعتين هما:

- المجموعة G_1: وهي مجموعة العناصر التي تقابل القيم($Y = 1$) [مجموعة الناجحين]، وتضم n_1 عنصراً، ويفترض أن يكون احتمال تحققها في كل تجربة ثابتاً ويساوي p، وإن p يقدر من $\tilde{p} = \frac{n_1}{n} = \frac{60}{100}$.

- المجموعة G_0: وهي مجموعة العناصر التي تقابل القيم $(Y = 0)$ [مجموعة الراسبين] وتضم n_0 عنصراً، ويفترض أن يكون احتمال تحققها في كل تجربة يكون ثابتاً ويساوي $q = 1 - p$.
ويقدر من: $\tilde{q} = \frac{n_0}{n} = \frac{40}{100}$.

وبناءً على نتائج هذه التجارب يمكننا تعريف عدة مؤشرات تستخدم في التحليل اللوجستي أهمها الأرجحية (*odds*) .

- **مفهوم الأرجحية (*odds*) وتعريفها:** يعود ظهور مفهوم الأرجحية إلى عمليات الرهان في الظواهر الثنائية (نجاح أو فشل , فوز أو خسارة) .

حيث يقال: إن إمكانية فوز اللاعب A تساوي n_1 مقابل n_0 للاعب المنافس $\bar{A}$. وإذا قام اللاعبان بإجراء $n = 100$ تجربة وفاز اللاعب A بـ $n_1 = 60$ تجربة وخسر $n_0 = 40$ تجربة منها، فإن تعريف الأرجحية لحادث فوز اللاعب A على اللاعب $\bar{A}$ يعطى بالعلاقة التالية:

$$odds(A) = \frac{n_1}{n_0} = \frac{\text{عدد مرات تحقق فوز } A}{\text{عدد مرات عدد فوز } \bar{A}} = \frac{60}{40} = \frac{3}{2} = \frac{1.5}{1} \qquad (2-9)$$

وعندها نقول أن أرجحية فوز اللاعب A على اللاعب $\bar{A}$ تساوي 60 مقابل 40 . وهنا يفضل اختصار الكسر $\frac{60}{40}$ إلى آخر عددين صحيحين مثل $\left(\frac{3}{2}\right)$، ونقول إن أرجحية فوز A على $\bar{A}$ تساوي 3 مقابل 2 . وإنه من الأفضل تحويل الكسر الأخير إلى نسبة عدد ما إلى الواحد مثل $\left(\frac{1.5}{1}\right)$، ونقول أن أرجحية فوز A على $\bar{A}$ تساوي 1.5 مقابل 1 . ونكتب ذلك على الشكل 1.5 : 1 .

- وبطريقة مشابهة نعرف الأرجحية لفوز اللاعب $\bar{A}$ بالعلاقة :

$$odds(\bar{A}) = \frac{n_0}{n_1} = \frac{\text{عدد مرات تحقق فوز } \bar{A}}{\text{عدد مرات عدد فوز } A} = \frac{40}{60} = \frac{2}{3} = \frac{1}{1.5} \qquad (3-9)$$

ونقول أن أرجحية فوز $\bar{A}$ على A تساوي 1 مقابل 1.5 . ونكتب ذلك على الشكل 1 : 1.5 .

- ومن التعريفين السابقين نستنتج أن:

$$odds(\bar{A}) = \frac{1}{odds(A)} \qquad (4-9)$$

$$odds(A) * odds(\bar{A}) = 1 \qquad (5-9)$$

- وأن احتمال تحقق فوز اللاعب (A) يعرف بالعلاقة التالية :

$$P(A) = \frac{n_1}{n_1 + n_0} = \frac{n_1}{n} = p = \frac{60}{100} = 0.60 = \frac{1.5}{1 + 1.5} \qquad (6-9)$$

- وأن احتمال تحقق فوز اللاعب $(\bar{A})$ يعرف بالعلاقة التالية :

$$P(\bar{A}) = \frac{n_1}{n_1 + n_0} = \frac{n_0}{n} = q = \frac{40}{100} = 0.40 = \frac{1}{1 + 1.5} \qquad (7-9)$$

ومن العلاقتين (9–6) و(9–7) نستخلص أن:

$$P(A) + P(\bar{A}) = p + q = 1 \qquad (8-9)$$

كما يمكننا استخلاص العلاقة التي ترتبط بين الأرجحية واحتمال تحقق حالتها، حيث نجد أنه يمكننا كتابة العلاقة (9–2) كما يلي:

$$odds(A) = \frac{n_1}{n_0} = \frac{\frac{n_1}{n}}{\frac{n_0}{n}} = \frac{P(A)}{P(\bar{A})} = \frac{p}{q} = \frac{p}{1-p} \qquad (9-9)$$

وكذلك نجد أن:

$$odds(\bar{A}) = \frac{n_0}{n_1} = \frac{\frac{n_0}{n}}{\frac{n_1}{n}} = \frac{P(\bar{A})}{P(A)} = \frac{q}{p} = \frac{1-p}{p} \qquad (10-9)$$

وسنستخدم العلاقة (9–9) في عمليات استخراج التابع اللوجستي [انظر الفقرة (9–6) في آخر هذا الفصل حول مفهوم الأرجحية وعلاقتها باحتمالات الظواهر الثنائية).

9–3: استخراج النموذج اللوجستي الثنائي:

لقد رأينا أن التابع Y (نتيجة الطالب في المثال (9–1)) هو تابع ثنائي ويأخذ إحدى القيمتين (1) للنجاح و(0) للرسوب، ومن شكل الانتشار (9–2) نلاحظ أن هذا التابع Y لا يصلح من وجهة نظر نظرية الانحدار، لأن يكون نتيجة لأي تابع خطي للمتحول المستقل X. لذلك يجب البحث عن بديل للتابع Y مرتبط به ويعبر عنه ويوصلنا إلى استخلاص علاقته مع X .

ولذلك نلجأ إلى مفهوم الاحتمال اللاحق لـ Y, وهو الاحتمال الشرطي لأن يأخذ التابع Y القيمة $(Y = 1)$ عندما تكون قيمة X محددة أو معلومة , وهو يساوي احتمال أن ينتمي العنصر المعلوم x إلى المجموعة G_1، ونكتب ذلك على الشكل التالي:

$$P(G_1/x) = P(Y = 1/x) = P_1(x) = [\text{الاحتمال اللاحق لـ } Y] \qquad (11-9)$$

وهو احتمال النجاح عند أية قيمة معلومة لـ x، وهو تابع مستمر ويأخذ قيمه في المجال [0 , 1]، وهو يصلح لأن يكون بديلاً عن Y، لأنه أصبح من الممكن رياضياً دراسة علاقة $P_1(x)$ مع المتحول المستقل x .

وإذا استطعنا أن نجد العلاقة بين هذا الاحتمال $P_1(x)$ والمتحول X، فإننا نكون قد تجاوزنا المشكلة، التي واجهتنا أثناء تمثيل Y الثنائي عبر X .

وهكذا نجد أنه يجب علينا الآن أن نقوم بإيجاد قيم $P_1(x)$ المقابلة لجميع قيم X، حتى نستطيع أن نقابلها مع قيم X، ثم استخلاص علاقة الانحدار بينهما دون وضع شروط مسبقة على المتحول X .

لذلك نقوم بحساب قيم الاحتمالات الشرطية $P_1(x)$ اللاحقة من علاقات (بايز), التي تأخذ الشكل التالي:

$$P_1(x) = P(G_1/x) = \frac{P * f(x/G_1)}{P * f(x/G_1) + q * f(x/G_0)} \qquad (12-9)$$

حيث أن: $f(x/G_1)$ و $f(x/G_0)$ هما التوزيعان التجريبيان لـ X ضمن المجموعتين G_1 وG_0 على الترتيب، وهما يحسبان (بعد تكرار التجربة n مرة) من التكرارات النسبية المقابلة لقيم X المختلفة كما يلي:

$$f(x/G_1) = \frac{n_1(x)}{n} \qquad (13-9)$$

$$f(x/G_0) = \frac{n_0(x)}{n} \qquad (14-9)$$

حيث أن: $n_1(x)$ هو عدد تكرار مرات النجاح مقابل القيمة (x) .

وأن: $n_0(x)$ هو عدد تكرار مرات الرسوب مقابل القيمة (x) .

وأن: $n_1 + n_0 = n$. وبتعويض ذلك في العلاقة(9−12) يمكننا حساب قيم الاحتمالات اللاحقة $P_1(x)$ المقابلة لمختلف قيم X ، ثم العمل على إيجاد العلاقة بينهما حسب قواعد نظرية الانحدار.

وسنحاول– في البداية– أن نفترض أن العلاقة بين $P_1(x)$ وX هي علاقة انحدار خطية من الشكل التالي :

$$\tilde{P}_1(x) = \propto + \beta x \qquad (15-9)$$

ثم نقوم بحساب تقدير لـ $\propto$ و β بطريقة المربعات الصغرى أو بطريقة الإمكانية العظمى، فنحصل على مستقيم انحدار محدد يفصل بين المجموعتينG_1 وG_0. كما هو مبين على الشكل (9−2) السابق.

ومنه نحسب القيم النظرية للاحتمالات اللاحقة $\tilde{P}_1(x)$ الواقعة على ذلك المستقيم مقابل كل قيمة لـ X .

ثم نقوم بحساب الاحتمالات اللاحقة المتممة له: $\tilde{P}_0(x)$ من العلاقة :

$$\tilde{P}_0(x) = 1 - \tilde{P}_1(x) \qquad (16-9)$$

وأخيراً نقوم بمقارنة $\tilde{P}_1(x)$ مع $\tilde{P}_0(x)$ ونصنف أي عنصر جديد x وفق القاعدة التالية:

(17 − 9) إذا $P_1(x) \geq P_0(x)$ نصنف x في المجموعة G_1 (في مجموعة الناجحين)

وإذا كان $P_1(x) < P_0(x)$ نصنفx في المجموعة G_0 (في مجموعة الراسبين)

والخط المستقيم على الشكل (9−2) يوضح ذلك .

ولكن الشكل (9−2) يظهر لنا أن جودة التمثيل لذلك المستقيم ضعيفة جداً (لأن قيمة R^2 صغيرة) . لذلك كان لابد من البحث عن حل آخر أو نموذج آخر لتمثيل العلاقة بين $P_1(x)$ وX، ومن أجل البحث عن تلك العلاقة، سنحاول الاستفادة من شكل العلاقة (9−11) ونستبدل $P_1(x)$ بتابع مستمر جديد ومناسب، وهو تابع الأرجحية (odds)، والذي يعرف بدلالة الاحتمال $P_1(x)$ من خلال العلاقة (9−9) والتي تأخذ الشكل التالي :

$$odds(x) = \frac{P_1(x)}{1 - P_1(x)} = \frac{\text{احتمال تحقق } Y}{\text{احتمال عدم تحقق } Y} = \frac{n_1}{n_0} \qquad (18-9)$$

حيث أن: $P_1(x)$ هو احتمال أن يأخذ التابع Y القيمة (1) عند القيمةx ، أو احتمال أن ينتمي العنصر x إلى المجموعة G_1، ونكتب ذلك كما يلي:

$$P_1(x) = P(Y = 1/x) = P(G_1/x) \qquad (19-9)$$

ولإيجاد علاقة الانحدار بين هذه الأرجحية (*odds*) والمتحول المستقل X، نفترض أنهما يرتبطان بعلاقة خطية لوغاريتمية كالعلاقة (9–4) السابقة، والتي نكتبها كما يلي:

$$ln(odds) = ln\left(\frac{P_1(x)}{1-P_1(x)}\right) = \propto + \beta x \qquad (20-9)$$

ويسمى التابع اللوغاريتمي الأيسر باسم $logit\big(P_1(x)\big)$ ويكتب على الشكل التالي :

$$logit[P_1(x)] = ln\left[\frac{P_1(x)}{1-P_1(x)}\right] = ln(odds) \qquad (21-9)$$

أي أن التابع $logit[P_1(x)]$ هو عبارة عن تحويل الاحتمال $P_1(x)$ المحسوب في (9–19) إلى (*odds*) ثم إلى الشكل اللوغاريتمي $ln\left[\frac{P_1(x)}{1-P_1(x)}\right]$، وهو عبارة عن تابع مستمر ويأخذ قيمه في المجال $]-\infty\,,+\infty[$ ، وذلك لأن $0 \le P_1(x) \le 1$: ، **فيكون** $0 \le \frac{P_1(x)}{1-P_1(x)} < +\infty$، **وبالتالي يكون لدينا أن:** $-\infty < ln\left[\frac{P_1(x)}{1-P_1(x)}\right] < +\infty$ ، **وهكذا نجد أنه يمكننا أن نفترض أن العلاقة بين التابع المستمر $logit[P_1(x)]$ والمتحول المستمر X هي علاقة خطية وتأخذ الشكل التالي:**

$$logit[P_1(x)] = \propto + \beta x = ln(odds) \qquad (22-9)$$

وبعد حساب القيم العددية لـ $logit[P_1(x)]$ من العلاقة (9–21) والعلاقات السابقة لها ، يمكننا إيجاد تقديرات لـ $\propto$ و β بتطبيق طريقة المربعات الصغرى أو طريقة الامكانية العظمى.

والآن نعود إلى العلاقة (9–20) فنجد أنه يمكننا إعادة كتابتها على الشكل التالي:

$$\frac{P_1(x)}{1-P_1(x)} = e^{\propto + \beta x} \qquad (23-9)$$

ومنها يمكننا أن نستخرج $P_1(x)$ كما يلي:

نقسم البسط والمقام في الطرف الأيسر على $P_1(x)$ فنحصل على أن:

$$\frac{1}{\frac{1}{P_1(x)} - 1} = e^{\propto + \beta x}$$

ثم نأخذ مقلوب الطرفين فنجد أن:

$$\frac{1}{P_1(x)} - 1 = \frac{1}{e^{\propto + \beta x}} = \bar{e}^{(\propto + \beta x)}$$

$$\frac{1}{P_1(x)} = 1 + \bar{e}^{(\propto + \beta x)}$$

ثم نأخذ مقلوب الطرفين مرة أخرى ونضرب البسط والمقام بالحد الأسي فنجد أن:

$$P_1(x) = \frac{1}{1+\bar{e}^{(\propto + \beta x)}} = \frac{e^{\propto + \beta x}}{1+e^{\propto + \beta x}} \qquad (24-9)$$

وبناء على (9–11) نجد أن الاحتمال اللاحق:

$$P(G_1/x) = P_1(x) = \frac{1}{1+\bar{e}^{(\propto + \beta x)}} = \frac{e^{\propto + \beta x}}{1+e^{\propto + \beta x}} \qquad (25-9)$$

وهو عبارة عن منحني التابع اللوجستي المرسوم على الشكل (9–2). وهنا نلاحظ أن هذا المنحني يختلف جذرياً عن المستقيم المرسوم على نفس الشكل، لأنه يقترب بطرفه الأيسر (في الأسفل) من نقاط المجموعة G_0، ويقترب بطرفه الأيمن(في الأعلى) من نقاط المجموعة G_1، وهو يعطينا بدقة أفضل، احتمال أن ينتمي x إلى G_1 مقابل كل قيمة x من قيم X .

ولحساب الاحتمال المتمم له نقوم بحساب $P_0(x)$ من العلاقة:

$$P_0(x) = 1 - P_1(x) = 1 - \frac{1}{1 + \bar{e}^{(\propto + \beta x)}} = \frac{\bar{e}^{(\propto + \beta x)}}{1 + \bar{e}^{(\propto + \beta x)}} \qquad (26-9)$$

وبتقسيم البسط والمقام على البسط نحصل على أن:

$$P_0(x) = P(G_0/x) = \frac{1}{1 + e^{\propto + \beta x}} \qquad (27-9)$$

قاعدة1: لاتخاذ قرار حول انتماء أي عنصر x لإحدى المجموعتين نطبق القاعدة التالية:

إذا كان $P_1(x) \geq P_0(x)$ نصنف x في المجموعة G_1

إذا كان $P_1(x) < P_0(x)$ نصنف x في المجموعة G_0 $\qquad (28-9)$

والمنحني المنطقي الملتوي على الشكل (9–2) يوضح ذلك .

ويمكننا تطوير أو تعديل القاعدة (9–28) السابقة لتصنيف العناصر x ، وذلك بأخذ نسبة الاحتمالين التالية: $\frac{P_1(x)}{P_0(x)}$, فنجد أن :

$$\frac{P_1(x)}{P_0(x)} = \frac{\frac{1}{1 + \bar{e}^{(\propto + \beta x)}}}{\frac{\bar{e}^{(\propto + \beta x)}}{1 + \bar{e}^{(\propto + \beta x)}}} = e^{\propto + \beta x} \qquad (29-9)$$

وبذلك تصبح قاعدة التصنيف لأي عنصر x كما يلي:

قاعدة2: نصنف أي عنصر x إلى المجموعة G_1 إذا كانت النسبة :

$$\frac{P_1(x)}{P_0(x)} \geq 1 \qquad \Leftrightarrow \qquad e^{\propto + \beta x} \geq 1 \qquad (30-9)$$

ونصنف x إلى المجموعة G_0 إذا كانت النسبة :

$$\frac{P_1(x)}{P_0(x)} < 1 \qquad \Leftrightarrow \qquad e^{\propto + \beta x} < 1 \qquad (31-9)$$

ويمكن تحويل هذه القاعدة إلى الشكل الخطي بأخذ اللوغاريتم الطبيعي للطرفين فنحصل على القاعدة التالية :

قاعدة3: نصنف أي عنصر x إلى المجموعة G_1 إذا كانت قيمة التركيب الخطي موجبة أو غير سالبة، أي إذا كان :

$$\propto + \beta x \geq 0 \qquad (32-9)$$

ونصنف أي عنصر x إلى المجموعة G_0 إذا كانت قيمته سالبة، أي إذا كان:

$$\propto +\beta x \quad < 0 \qquad\qquad (33-9)$$

وبذلك نحصل على توابع تمييزية فاصلة بين المجموعات بأساليب متعددة. ولكن الاختلاف بين هذه القواعد والقواعد الخطية المستخدمة في التحليل التمييزي، هو أنها تستند على النسبة بين التوزيعات الاحتمالية اللاحقة للمجموعات, بينما تعتمد القواعد المستخدمة في التحليل التمييزي على النسبة بين الاحتمالات السابقة P_2 و P_1 .

كما إن شكل العلاقة (9–25) لـ $P_1(x)$ يساعدنا في الحصول على تقدير لـ $\propto$ و β بطريقة الامكانية العظمى.

9–4–: تقدير معالم النموذج اللوجستي (بطريقة الامكانية العظمى MLE) بمتحول واحد:

(*Maximum Likelihood Estimation*) *X* .*[Webb P.159* بتصرف وإضافة] .

إن تقدير معالم النموذج اللوجستي $\propto$ و β بطريقة الإمكانية العظمى *MLE* يعتمد على بيانات إحصائية معينة، ويتطلب حساب المواصفات الميدانية لتلك للبيانات، ويرتبط بتصميم المعاينة التي تطبق على المجموعتين G_1 و G_0 .

وهناك عدة تصاميم لهذه المعاينة هي:

1– المعاينة المختلطة : وتكون من توزيعات مختلطة بين المجموعتين G_1 و G_0 (توزيع مشترك) .

2– المعاينة الشرطية لـ X: تجري بحيث يكون x ثابتاً، ثم نسحب عينة أو أكثر من العناصر (التي يمكن أن تنتمي إلى G_1 أو إلى G_0).

3– المعاينة المنفصلة من كل مجموعة على حدة : حيث تكون التوزيعات الشرطية $P(x/G_1)$ أو $P(x/G_0)$ هي التوزيعات المعتمدة.

علماً بأن طريقة الإمكانية العظمى تعطينا تقديرات للمعلم β، تكون مستقلة عن شكل تصميم المعاينة. وإن بعض تصاميم المعاينة تعطينا تقديرات أفضل من تصاميم أخرى للمعلم β_0 (تصميم المعاينة المنفصلة).

والآن لنفترض إننا نعمل ضمن المعاينة المختلطة، التي تفترض أن العينة العشوائية مسحوبة من المجتمع المختلط للمجموعتين بحجم *n*، ومؤلفة من: n_1 عنصراً من G_1 ,و n_0 عنصراً من المجموعة G_0، والتي سنرمز لها لضرورات رياضية بالرمز G_2 ولعدد عناصرها بـ n_2، وعندها نجد أن تابع الإمكانية العظمى L في هاتين المجموعتين يأخذ شكل الجداء التالي:

$$L = L_1 * L_2 = \prod_{i=1}^{n_1} P(x_{1i}/G_1) * \prod_{i=1}^{n_2} P(x_{2i}/G_2) \qquad (34-9)$$

حيث أن n_s ... $i: 1\ 2\ 3$ وحيث أن: x_{si} هي المشاهدة المسحوبة من المجموعة G_s وأن: $s: 1\ 2$,

وبما أن التوزيع الشرطي لـ $P(x\ /G_s)$ يساوي :

$$P(x\ /G_s) = \frac{P(x) * P(G_s/x)}{P(G_s)} \qquad : s = 1\ 2 \qquad (35-9)$$

نقوم الآن بتعويض $P(x/G_s)$ من (9–35) في العلاقة (9–34) فنحصل على أن:

$$L = \prod_{i=1}^{n_1} P(G_1/x_{1i}) \frac{P(x_{1i})}{P(G_1)} * \prod_{i=1}^{n_2} P(G_2/x_{2i}) \frac{P(x_{2i})}{P(G_2)} \qquad (36-9)$$

وبإخراج التوزيعين $P(G_1)$ و$P(G_2)$ خارج الجداء لأنه ليس لهما علاقة بدليل الجداء i ، نجد أن:

$$L = \frac{1}{P(G_1) * P(G_2)} * \prod_{i=1}^{n_1} P(x_{1i}) * P(G_1/x_{1i}) * \prod_{i=1}^{n_2} P(x_{2i}) * P(G_2/x_{2i}) \quad \mathbf{(37-9)}$$

وبما أن $P(x_{1i})$ و$P(x_{2i})$ ليس لهما علاقة بالمعلمتين ∝ و β للنموذج، لذلك نكتب جداءاتهما على الشكل التالي:

$$\prod_{i=1}^{n_1} P(x_{1i}) * \prod_{i=1}^{n_2} P(x_{2i}) = \prod_{i=1}^{n} P(x_i) : \quad \mathbf{(38-9)}$$

وهنا نلاحظ أن الجداء الأخير قد أصبح مأخوذاً على كامل حجم العينة $\boldsymbol{n}$

وبذلك نجد بأن تابع الامكانية العظمى يأخذ الشكل التالي:

$$L = \frac{\prod_{i=1}^{n} P(x_i)}{P(G_1) * P(G_2)} * \prod_{i=1}^{n_1} P(G_1/x_{1i}) * \prod_{i=1}^{n_2} P(G_2/x_{2i}) \qquad (39-9)$$

وبما أن الحد الأول $\frac{\prod P(x_i)}{P(G_1)*P(G_2)}$ ليس له علاقة بالمعلمتين ∝ و β للنموذج اللوجستي . لذلك يمكننا افتراض (كما فعل اندرسون 1967), أن التابع L مستقل عن الاحتمالات السابقة $P(x)$. وبذلك يمكننا اختصار التابع L إلى تابع مكافئ له L' يساوي:

$$L' = \prod_{i=1}^{n_1} P(G_1/x_{1i}) * \prod_{i=1}^{n_2} P(G_2/x_{2i}) \qquad (40-9)$$

والآن نأخذ اللوغاريتم الطبيعي للطرفين في (9–40) فنجد أن:

$$ln\, L' = \sum_{i=1}^{n_1} ln\, P(G_1/x_{1i}) + \sum_{i=1}^{n_2} ln\, P(G_2/x_{2i})$$

وبتعويض ما تحت اللوغاريتمات بما تساويها من العلاقتين (9–25) و(9–27)، نحصل على أن $ln\, L'$ يساوي:

$$ln\, L' = \sum_{i=1}^{n_1} (\propto + \beta X_{1i}) - \sum_{i=1}^{n_1} ln\big(1 + e^{\propto + \beta x_{1i}}\big) - \sum_{i=1}^{n_2} ln\big(1 + e^{\propto + \beta x_{2i}}\big) \quad \mathbf{(41-9)}$$

وبعد دمج المجموعين الأخيرين وأخذ المجموع على كامل العينة n نجد أن:

$$ln\, L' = \sum_{i=1}^{n_1} (\propto + \beta X_{1i}) - \sum_{i=1}^{n} ln\big(1 + e^{\propto + \beta x_i}\big) \qquad (42-9)$$

والآن نقوم بأخذ المشتقين الجزئيين لـ($\ell n\, L'$) بالنسبة لـ ∝ و β ونضعهما مساويين للصفر، فنحصل بعد الإصلاح والاستبدال على أن هذين المشتقين يساويان:

$$\frac{\partial\, \ell n\, L'}{\partial \propto} = n_1 - \sum_{i=1}^{n} P(G_1/x_i) = 0 \qquad (43-9)$$

$$\frac{\partial\, \ell n\, L'}{\partial \beta} = \sum_{i=1}^{n_1} x_{1i} - \sum_{i=1}^{n} (x_i) * P(G_1/x_i) = 0 \qquad (44-9)$$

مع الانتباه إلى أن المجموعين الأخيرين ($\sum_{i=1}^{n}$(مأخوذين على جميع قيم x في العينة n، ثم نقوم بحل هاتين المعادلتين فنحصل على تقدير لـ ∝ وβ، ومنهما نحصل على النموذج اللوجستي المطلوب.

ملاحظة: إذا كان عدد المتحولات المؤثرة X يساوي p متحولاً فإننا سنرمز لها بشعاع واحد كما يلي:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$$

وعندها يمكننا كتابة العلاقة (9−20) كما يلي:

$$ln\left[\frac{P_1(x)}{1-P_1(x)}\right] = \propto + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_p X_p = \propto + \beta' X \qquad (45-9)$$

حيث أن: $\beta'(\beta_1\ , \beta_2 \dots \beta_p\ )$

وعندها أيضاً فإن صيغة النموذج اللوجستي تأخذ الشكل التالي:

$$P_1(x) = \frac{1}{1+\bar{e}^{(\propto+\beta' X)}} = \frac{e^{(\propto+\beta' X)}}{1+e^{(\propto+\beta' X)}} \qquad (46-9)$$

ثم نجد أن $P_0(x)$ تحسب من العلاقة:

$$P_0(x) = 1 - P_1(x) = \frac{1}{1+e^{\propto+\beta_1' X}} \qquad (47-9)$$

وعندها فإن معادلات الإمكانية العظمى لحساب ∝ وβ' تأخذ الشكل التالي:

$$\frac{\partial\, ln\, L'}{\partial \propto} = n_1 - \sum_{i=1}^{n} P(G_1/x_i) = 0 \qquad (48-9)$$

$$\frac{\partial\, ln\, L'}{\partial \beta_j} = \sum_{i=1}^{n_1} (x_{1i})_j - \sum_{i=1}^{n} (x_i)_j * P(G_1/x_i)_j = 0 \qquad (49-9)$$

حيث أن: J تأخذ القيم $j: 1\ 2\ 3 \dots P$

ومن هذه المعادلات يمكننا حساب تقديرات لـ ∝ و$\beta_1\ , \beta_2 \dots \beta_p$، علماً بأن المعادلات (9−48) و(9−49) هي معادلات غير خطية بالنسبة لـ ∝ وβ، وإن حلها يحتاج إلى أساليب وبرامج متقدمة.

مثال (9–2): بناءً على بيانات المثال (9–1) السابق تم تقدير معالم النموذج (9–24) بطريقة الامكانية العظمى، فكانت النتيجة كما في الجدول التالي: [المصدر: *Wikipedia*] :

جدول (9–3): نتائج علاقة الانحدار لـ *logit P(X)* على *X*

البيان	قيم الأمثال	الانحراف المعياري	قيمة *Z* *Wald*	قيمة *P* حسب اختبار *(Wald)*
الثابت	$\propto = -4.0777$	1.7610	−2.316	0.0206
الساعات *X*	$\beta = 1.5046$	0.6287	2.393	0.0167

نلاحظ أن هذه المخرجات تشير إلى أن عدد ساعات الدراسة *X* ترتبط معنوياً مع احتمال النجاح في الامتحان (وذلك لأن قيمة *P* حسب اختبار *Wald* تساوي *P = 0.0167*، وهي أصغر من قيمة مستوى الدلالة 0.05).

وإن معادلة العلاقة بين $logit[P_1(x)]$ و*X* تساوي :

$$logit[P_1(x)] = ln\left[\frac{P_1(x)}{1-P_1(x)}\right] = 1.5046X - 4.0777 = 1.5046[X - 2.71]$$

وهكذا نجد أن معادلة الأرجحية مع *X* هي:

$$odds(Y) = \frac{P_1(x)}{1-P_1(x)} = e^{1.5046(X-2.71)}$$

ومنها نجد أنه إذا كان عدد الساعات *X = 2.71* فإن:$odds(Y) = e^0 = 1$، وفي هذه الحالة نستنتج أن إمكانية النجاح ($odds$) تساوي إمكانية الرسوب، وإن احتمال كل منهما يساوي 0.50، وفي هذه الحالة (عندما *X=2.71*) يمكننا أن نقول أن : أرجحية النجاح مقابل الرسوب هي كما يلي: واحد مقابل واحد ونكتب ذلك كما يلي (1 : 1).

ولحساب الاحتمال $P_1(x)$ من النموذج السابق نعتمد على العلاقة (9–46) فنجد أن:

$$P_1(x) = \frac{1}{1+\bar{e}^{(\tilde{\propto}+\tilde{\beta}X)}} = \frac{1}{1+\bar{e}^{(-4.0777+1.5046X)}}$$

فعندما يكون *X = 2* نجد أن احتمال النجاح يساوي :

$$P_1(2) = \frac{1}{1+\bar{e}^{(-4.0777+1.5046*2)}} = 0.26 \quad\Longrightarrow\quad P_0(2) = 0.74$$

وهذا يعني أن احتمال نجاح من يدرس ساعتين (*X=2*) يساوي(0.26) وهو أصغر من احتمال رسوبه (0.74)، لذلك نصنف ذلك الطالب في المجموعة G_0 ونعتبره من مجموعة الراسبين .

أما عندما يكون *X=2.71* فإننا نجد أن احتمال النجاح يساوي:

$$P_1(x) = \frac{1}{1+e^0} = \frac{1}{2} = 0.50$$

وهذا ما لاحظناه سابقاً وهذا يعني أن القيمة X=2,71 هي النقطة الفاصلة بين النجاح والرسوب .

أما عندما تكون *X=4* فإن احتمال النجاح $P_1(x)$ يساوي :

$$P_1(4) = \frac{1}{1+\bar{e}^{(-4.0777+1.5046*4)}} = 0.87 \quad \Rightarrow \quad P_0(4) = 0.13$$

وهذا يعني أن احتمال نجاح من يدرس (*X=4*) ساعات يساوي(0,87) وهو أكبر من احتمال رسوبه (0.13)، لذلك نصنف ذلك الطالب في المجموعة G_1 [مجموعة الناجحين] .

أما عندما يكون *X=3* فإن احتمال النجاح يساوي :

$$P_1(3) = \frac{1}{1+\bar{e}^{(-4.0777+1.5046X)}} = 0.61 \quad \Rightarrow \quad P_0(3) = 0.39$$

أي أن احتمال نجاح من كان يدرس (*X=3*) ساعات يساوي(0.61) وهو أكبر من احتمال رسوبه (0.39)، ولذلك نصنف ذلك الطالب في المجموعة G_1 [الناجحين]، رغم إنه كان من بين الراسبين (انظر الجدول (6−1)).

وأخيراً يمكننا أن ننظم بعض النتائج الممكنة لهذا النموذج في جدول مناسب كالجدول التالي :

جدول (9−4): قيم التحليل اللوجستي

احتمال النجاح $P_1(x)$	قيمة الـ *odds*	قيمة*ln(odds)*	عدد ساعات الدراسة *X*
0.07	$0.078 \approx 1:13.1$	−2.57	1
0.26	0.034 $\approx 1:12.91$	−1.07	2
0.50	$1 \approx 1:1$	0.00	2.71
0.61	$1.55 \approx \cdots$	0.44	3
0.87	$6.96 \approx \cdots$	1.94	4
0.97	$31.4 \approx \cdots$	3.45	5
0.99	141.16085	4.95	6

وبالعودة إلى الجدول (9−3) نجد أن احتمال الدلالة يساوي $P = 0.0167$، علماً بأن هذه القيمة محسوبة استناداً إلى علاقة اختبار *Wald −Z*. ولكن هناك طريقة أفضل من طريقة *Wald* − الطريقة المعتمدة في حساب قيمة *P* للتوابع اللوجستية − وهي طريقة اختبار نسبة الامكانية العظمى (*LRT*)، والتي تعطينا من هذه البيانات أن قيمة *P* للنموذج اللوجستي المدروس $P = 0.0006$.

9−5− التحليل اللوجستي المتعدد: [webb P. 161 بتصرف]

إن التحليل اللوجستي المتعدد هو تعميم للتحليل اللوجستي الثنائي، وهو يعالج الحالات التي يكون فيها المجتمع مؤلفاً من عدة مجموعات (أو فئات) منفصلة نرمز لها بـ :

$$G_1 \; G_2 \; \dots \; G_j \; \dots \; G_g \qquad (50-9)$$

وعندها فإننا نشكل تابع الـ *logit* لكل زوج من هذه المجموعات، وإذا أخذنا المجموعة G_j مقابل أية مجموعة أخرى ولتكن G_g ، فإن تابع الـ *logit* يأخذ الشكل التالي:

$$logit\,[P_j(x)] = ln\left[\frac{P_j(x)}{P_g(x)}\right] = \propto_j + \beta'_j X \qquad (51-9)$$

حيث أن: $j = 1\ 2\ 3 \ldots g-1$ وأن: $X(X_1\ X_2\ \ldots\ X_P)$ وأن: $\beta'(\beta_1\ \beta_2\ \ldots\ \beta_P)$

وهي تمثل $(g-1)$ تابعاً تمييزياً تفصل بين تلك المجموعات.

وهذا يعني أن لوغاريتم نسبة الإمكانية، $\left[\frac{P_j(x)}{P_g(x)}\right]$ لأي زوج ممكن من المجموعات يرتبط مع المتحولات X بعلاقة خطية، وهي تشكل المستوى الفاصل بينهما .

وبطريقة مشابهة لما عرضناه في التحليل اللوجستي الثنائي يمكننا صياغة الاحتمالات اللاحقة $P_j(x)$ و $P_g(x)$ بدلالة X بواسطة العلاقتين التاليتين :

$$P_j(x) = P(G_j/x) = \frac{e^{\propto_j + \beta'_j X}}{1 + \sum_{j=1}^{g-1} e^{\propto_j + \beta'_j X}} \qquad (52-9)$$

حيث أن: $j = 1\ 2\ 3 \ldots g-1$ وأن: $(g-1)$ هو عدد التوابع في (51-9), أما التابع المتمم $P_g(x)$ فيساوي :

$$P_g(x) = P(G_g/x) = \frac{1}{1 + \sum_{j=1}^{g-1} e^{\propto_j + \beta'_j X}} \qquad (53-9)$$

وهكذا نجد أن قاعدة التصنيف التمييزي تصبح تابعة للعلاقة الخطية $(\propto_j + \beta'_j X)$. وتأخذ الصيغة التالية:

قاعدة: نصنف x إلى المجموعة G_k إذا كانت $\propto_k + \beta'_k x \ > 0$ وكانت قيمته أكبر من القيم الأخرى: أي إذا كانت:

$$0 < \propto_k + \beta'_k x \ = Max[\propto_j + \beta'_j X] \qquad (54-9)$$

حيث أن: $j = 1\ 2\ 3 \ldots g-1$, أما إذا كان العكس فنصنفه إلى المجموعة G_g .

وكذلك يمكننا أن نستخلص تابع الإمكانية العظمى من العلاقة:

$$L = \prod_{j=1}^{g} \prod_{i=1}^{n_i} P(x_{ji}/G_j) \qquad (\mathbf{55-9})$$

وبإجراء نفس العمليات والمعالجات على L نحصل على التابع المكافئ له L' ونأخذ لوغاريتمه فنجد أن:

$$ln(L') = \sum_{j=1}^{g} \sum_{i=1}^{n_j} ln\, P(G_j/x_{ji}) \qquad (56-9)$$

وباشتقاقه نحصل على المعادلات التالية:

$$\frac{\partial ln(L')}{\partial \propto} = n_j - \sum_{X\text{لك}}^{n} P(G_j/X) = 0 \qquad (57-9)$$

$$\frac{\partial ln(L')}{\partial(\beta_j)_s} = \sum_{i=1}^{n_j}(x_{ji})_s - \sum_{X\text{لك}}^{n} x_s * P(G_j/X)_s = 0 \qquad \mathbf{(58-9)}$$

حيث أن: *j: 1 2 3 ..... g−1*

وبحل هذه المعادلات نحصل على تقديرات للمعلمات $\propto$ وβ ، ولكن بما أن هذه المعادلات ليست خطية، فإنه يتم البحث عن حلول تقاربية لها (بطريقة نيوتن أو بطريقة المعاودة)، وعندها نحتاج إلى حل ابتدائي ننطلق منه لإيجاد الحلول المتتالية والمتقاربة، ويمكن أن نأخذ الحل الابتدائي المقابل للقيم الصفرية، ونضع في البداية $\propto_j= 0$ و$\beta_1 = \beta_2 = \cdots \beta_g = 0$، ثم نتابع البحث عن قيم $\propto$ وβ المثالية.

6-9: تقييم جودة النموذج اللوجستي: [Wikipedia بتصرف]:

إن عملية تقييم جودة النموذج اللوجستي تختلف عن عملية تقييم الجودة في الانحدار الخطي، ومع أن الانحدار اللوجستي يقدر معالم النموذج β_j بطريقة الإمكانية العظمى، فهو لا يعتمد على معامل التحديد R^2 المعروف لتقييم جودة التمثيل . ولكنه يعتمد في تقييم جودة التمثيل على مفاهيم جديدة هي:

- **النموذج المشبع (Saturated Model)**: وهو النموذج الذي يمثل (نظرياً) البيانات المدروسة تمثيلاً تاماً، ونرمز لتابع الإمكانية العظمى لهذا النموذج بالرمز L_S . مأخوذة من (Likelihood of the saturated model) ولكن عملية الحصول على هذا النموذج في الحالة العامة قد تكون غير ممكنة . ويبقى L_S مجهولاً.
- **النموذج الصفري (Null Model)**: وهو النموذج الذي يتوافق مع فرضية العدم $(H_0: \beta_j = 0)$. أي أنه النموذج الذي لا يتضمن أي من المتحولات X، ويأخذ قيمة ثابتة هي قيمة الثابت β_0 أو$\propto$. ونرمز لتابع الإمكانية لهذا النموذج بالرمز L_0 .
- **النموذج المقدر (Fitted Model)**: وهو النموذج الذي ينتج عن حساب وتقدير المعالم β_i بطريقة الإمكانية العظمى، وهو يتضمن بعض المتحولات X (واحد على الأقل). ونرمز لتابع الإمكانية العظمى لهذا النموذج بالرمز L_M . علماً بأن توابع الإمكانية العظمى L_0 وL_M تحسب اعتماداً على العلاقة (9−34) أو على العلاقة (9−40) وهي تأخذ شكل جداءات توزيع (بيرنوبللي) التالي:

$$L = \prod_{\substack{i=1 \\ j=1,2}}^{n} P(G_j/x_{ji}) = \prod_{i=1}^{n} \left(P_1(x_i)\right)^{y_i}[1 - P(x_i)]^{1-y_i} \qquad (59-9)$$

حيث y_i تأخذ (1) أو (0) .

وبناء على ذلك تم تعريف المؤشرات التالية:

1- **حيدان النموذج المقدر (Deviance Model)**: ويعرف بالعلاقة التالية:

$$D_{\not{f}} = -2\ln\frac{(\text{قيمة تابع الإمكانية العظمى للنموذج المقدر})}{(\text{قيمة تابع الإمكانية العظمى للنموذج المشبع})} = -2\ln\left[\frac{L_M}{L_S}\right] \quad (60-9)$$

وهو يعبر عن الاختلاف النسبي بين النموذج (بمتحول واحد أو أكثر) وبين النموذج المشبع، لقد تم وضع الإشارة السالبة قبل اللوغاريتم لأن $L_M < L_S$.

2- **حيدان النموذج الصفري**: ويعرف كما يلي:

$$\boldsymbol{D_0 = -2\ln\frac{(\text{قيمة تابع الإمكانية العظمى للنموذج الصفرى})}{(\text{قيمة تابع الإمكانية العظمى للنموذج المشبع})} = -2\ln\left[\frac{L_0}{L_S}\right] \quad (61-9)}$$

3- **الفرق** Δ : ويحسب للتخلص من L_S المجهولة، وعند حساب الفرق بين D_0 و $D_{\not{f}}$ نجد أن:

$$\Delta = D_0 - D_{\not{f}} = -2\,ln\left[\frac{L_0}{L_S}\right] + 2\,ln\left[\frac{L_M}{L_S}\right] = -2\left[ln\left(\frac{L_0}{L_S}\right) - ln\left(\frac{L_M}{L_S}\right)\right]$$

$$\Delta = D_0 - D_{\not{f}} = -2\ln\left[\frac{\frac{L_0}{L_S}}{\frac{L_M}{L_S}}\right] = -2\,ln\left[\frac{L_0}{L_M}\right] = -2[ln\,L_0 - ln\,L_M\,] \qquad (62-9)$$

ويستخدم هذا الفرق Δ في تقييم جودة التمثيل، لأنه من هذه العلاقات يمكننا أن نستنتج أن: $L_0 < L_M < L_S$، وأن: $D_0 > D_{\not{f}}$ ، وأنه كلما ازدادت قيمة L_M واقتربت من L_S كان التمثيل جيداً، ولكن زيادة L_M تعني زيادة $D_{\not{f}}$، وهذا يؤدي إلى تناقص الفرق $\Delta = D_0 - D_{\not{f}}$، وهذا يعني أنه كلما تناقص الفرق Δ كان التمثيل جيداً . لذلك يجب علينا عند تقدير المعالم ∝ وβ البحث عن النموذج الذي يجعل Δ أصغر ما يمكن، ولهذا السبب تخصص البرامج الحاسوبية عموداً خاصاً للفرق $\Delta = -2\,ln\left[\frac{L_0}{L_M}\right]$ ، وتراقب تغيراته وتتخذه كمعيار للتوصل إلى الحل المثالي للمعادلات (9-48) و(9-49) بطريقة المعاودة، وذلك حتى تستقر قيمة الفرق Δ عند قيمة ثابتة خلال المعاودات الأخيرة.

4- **دراسة قيم معاملات التحديد اللوجستية**: هناك عدة معاملات تحديد لتقييم جودة النموذج اللوجستي (أكثر من عشرة)، ولكن أهمها المعاملات التالية :

- معامل نسبة الإمكانية العظمى (Likelihood Ratio):

$$R_L^2 = \frac{D_0 - D_{\not{f}}}{D_0} = 1 - \frac{ln\left[\frac{L_M}{L_S}\right]}{ln\left[\frac{L_0}{L_S}\right]} \qquad (63-9)$$

- معامل cox & snele:

$$R_{cs}^2 = 1 - \left(\frac{L_0}{L_M}\right)^{\frac{2}{n}} = 1 - e^{\frac{2[\ln L_0 - \ln L_M]}{n}} \qquad (64-9)$$

- معامل Nagelkerke:

$$R_N^2 = \frac{1 - \left(\frac{L_0}{L_M}\right)^{\frac{2}{n}}}{1 - (L_0)^{\frac{2}{n}}} \quad : \quad 0 < R_N^2 < 1 \qquad (65-9)$$

- معامل Mc Fadden:

$$R_{MCF}^2 = 1 - \frac{\ln(L_M)}{\ln(L_0)} \qquad (66-9)$$

وإن هذه المعاملات ترتبط مع بعضها بالعلاقات التالية :

$$R_{cs}^2 = 1 - \left(\frac{1}{L_0}\right)^{\frac{2}{n}R_{MCF}^2} \qquad (67-9)$$

$$R_N^2 = \frac{R_{cs}^2}{1 - (L_0)^{\frac{2}{n}}} \qquad (68-9)$$

$$R_{MCF}^2 = \frac{-n}{2} * \frac{\ln(1 - R_{cs}^2)}{\ln(L_0)} \qquad (69-9)$$

5- **حساب جدول تقاطع حالات التصنيف اللوجستي مع التصنيف الفعلي** (السابق): والذي يأخذ الشكل التالي (في حالة مجموعتين) :

جدول (9-5):تكرارات التقاطعات

اللوجستي / الفعلي	G_1	$G_2 = G_0$	المجموع
G_1	n_{11}	n_{12}	n'_1
$G_2 = G_0$	n_{21}	n_{22}	$n'_2 = n_0$
المجموع	n_1	n_2	n

ومنه يتم حساب معدل التصنيف الصحيح بحساب نسبة مجموع عناصر القطر الرئيسي على المجموع الكلي n فنجد أن:

$$R = \frac{n_{11} + n_{22}}{n} 100\% \qquad (70-9)$$

ومنه يمكن حساب معدل التصنيف الخاطئ :

$$MR = 1 - R = \frac{n_{12} + n_{21}}{n} 100\% \qquad (71-9)$$

6- **حساب المعامل** kappa من العلاقة:

$$kappa = \frac{n\sum^2 n_{ii} - \sum^2 n_i * n_i'}{n^2 - \sum^2 n_i * n_i'} 100\% \qquad (72-9)$$

وهناك مقاييس أخرى لتحليل الجودة مثل هذه الجداول، كالحساسية (SE) والخصوصية (SP) ونسبة الأرجحية (OR) وغيرها، وهي مذكورة ومعرفة في العلاقات (9–79) و(9–81) و(9–83) من الفقرة اللاحقة (9–7) .

7– **اختبار (هوسمير– ليمشو Hosmer– Lemshow): لجودة المطابقة:** وهو يختبر صحة فرضية العدم التالية H_0: يتساوى عدد الحالات المشاهدة مع عدد الحالات المتوقعة (أي أن النموذج يمثل البيانات بشكل صحيح). وعندها نتخذ القرار بقبول فرضية العدم H_0، إذا كان مستوى المعنوية أو احتمال الدلالة p لاختبار (كاي مربع) أكبر من مستوى الدلالة المحدد بـ ∝ .

8– **دراسة معنوية المتحولات الداخلة في النموذج:** وذلك من خلال مقارنة قيم احتمال الدلالة $(P = sig)$ مع مستوى الدلالة $\propto = 0.05$. فإذا كانت قيمة P_j المقابلة للمتحولات X_j أصغر من 0.05 ، تكون علاقة ذلك المتحول مع النموذج معنوية والعكس بالعكس . ولكن دراسة هذه المعنوية تعتمد على اختبار جديد هو اختبار (*Wald*) وهو اختبار شبيه بالاختبار الطبيعي ويعرف بالعلاقة التالية: $Wald = \left(\frac{\beta_i}{SE(\beta_i)}\right)^2$.

9– **دراسة قوة تأثير كل متحول X_j على النموذج:** وذلك من خلال حساب نسبة الأرجحية له OR (Odds Ratio) والتي يمكن تعريفها وحسابها من العلاقة التالية :

$$\mathrm{OR} = \frac{odds(x_j + 1)}{odds(x_j)} = \frac{e^{\tilde{\propto} + \tilde{\beta}(x_j+1)}}{e^{\tilde{\propto} + \tilde{\beta}(x_j)}} = e^{\beta} \qquad (73-9)$$

لذلك يقوم البرنامج الحاسوبي بحساب القوة e^{β} لجميع المتحولات الداخلة في النموذج ويضعها في عمود خاص، لأنها تدل على قوة تأثير المتحول X_j عندما يزداد بمقدار واحد (1) على النموذج ككل، وهي تعبر عن مرونة المتحول X_j بالنسبة للتابع (odds) .

مثال (9–3): في دراسة نشرها [شاهين 2014] حول تطبيق التحليل اللوجستي على مرضى سرطان الدم (اللوكيميا) في محافظة البصرة، لاحظ أولاً أن الجهات الطبية تصنف هؤلاء المرضى إلى نوعين أو مجموعتين هما: مرض سرطان الدم النخاعي الحاد (AME) ومرض سرطان الدم اللمفاوي (ALL).

لذلك قام الباحث بسحب عينيتين عشوائيتين بحجمين متساويين $n_1 = n_2 = 80$ من ملفات هاتين المجموعتين، ثم قام بدراستها وتحليلها وتحديد المتحولات التي تفسر عملية الإصابة بهذين المرضين، فكانت ثمانية متحولات مختلطة (كمية ونوعية)، ثم قام بتحويلها إلى متحولات ثنائية وصنف قيم كل منها ضمن فئتين محددتين كما يلي:

X_1 – جنس المريض وصنفه إلى (ذكر = 1 ، انثى = 2) .

X_2 – عمر المريض وصنفه إلى [$(X_2 \leq 50) = 1$, $(X_2 > 50) = 2$].

X_3 – وزن المريض وصنفه إلى [$(X_3 \leq 40) = 1$, $(X_3 > 40) = 2$].

X_4 – نسبة الكريات الحمراء (P,C,V) وصنفه إلى [$(X_4 \leq 0.35) = 1$, $(X_4 > 0.35) = 2$].

X_5 – هيموغلوبين الدم (H,B) وصنفه إلى [$(X_5 \leq 11.5) = 1$, $(X_5 > 11.5) = 2$].

X_6 – معدل الكريات البيضاء (W,B,C) وصنفه إلى [$(X_6 \leq 7.5) = 1$, $(X_6 > 7.5) = 2$].

X_7– سرعة ترسب الكريات الحمراء (E,S,Q) وصنفه إلى [$(X_7 \leq 22.5) = 1$, $(X_7 > 22.5) = 2$]

X_8 – عدد الصفائح الدموية (P,C) وصنفه إلى [$(X_8 \leq 150) = 1$, $(X_8 > 150) = 2$].

كما قام بتسمية تابع الاستجابة الثنائي Y، الذي يعبر عن نوع الإصابة بمرض سرطان الدم وصنف حالتيه (النخاعي واللمفاوي) كما يلي:

إذا كان المريض مصاب بسرطان الدم النخاعي فإن قيمة $Y = 0$.

إذا كان المريض مصاب بسرطان الدم اللمفاوي فإن قيمة $Y = 1$.

ثم قام بجمع البيانات اللازمة له من ملفات عناصر العينة المسحوبة من المجموعتين، وعمل على تحليلها وتحويلها إلى متحولات ثنائية، ووضعها في جداول منظمة، ثم قام بتطبيق برنامج التحليل اللوجستي على المتحولات الثنائية التالية :

$$Y: X_1 , X_2 , X_3 , X_4 , X_5 , X_6 , X_7 , X_8$$

واستخدم لذلك طريقة (enter)، واختار طريقة (نيوتن) لإجراء المعاودة لإيجاد الحلول التقريبية للمعادلات غير الخطية الواردة في العلاقات (9–48) و(9–49)...الخ . وذلك من أجل الحصول على معالم النموذج اللوجستي التالي :

$$\ln\left(\frac{P_1(x)}{1-P_1(x)}\right) = \propto + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_8 X_8$$

فكانت النتائج بعد إجراء (6) دورات للمعاودة كما يلي:

جدول (9–6): قيم الأمثال للتابع اللوجستي خلال دورات المعاودة:

رقم الدورة	$\Delta = -2\,\ell n\left[\frac{L_0}{L_M}\right]$	الثابت C	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	162.878	0.758	−2.268	−0.207	−0.871	−1.080	0.557	−0.359	−0.195	0.949
2	156984	0.901	−3.387	−0.236	−1.184	−1.370	0.669	−0.474	−0.194	1.246
3	156.243	0.913	−3.998	−0.228	−1.258	−1.130	0.679	−0.497	−0.185	1.212
4	156.200	0.912	−4.188	−0.227	−1.264	−1.416	0.679	−0.499	−0.183	1.316
5	156.200	0.912	−4.206	−0.227	−1.264	−1.416	0.679	−0.499	−0.183	1.316
6	156.200	0.912	−4.206	−0.227	−1.264	−1.416	0.679	−0.499	−0.183	1.316

وهنا نلاحظ أن قيم أمثال التابع اللوجستي قد استقرت تماماً بعد إجراء خمس أو ست دورات لمعاودة الحسابات

كما إن قيمة مؤشر الجودة للنموذج المتمثل بالعمود الثاني الذي يتضمن قيم الفرق $(D_0 - D_f)$، والتي استقرت عند القيمة 156,200 في الدورة الرابعة ومابعدها. لذلك توقفت الحسابات عند الدورة السادسة.

ثم انتقل الباحث إلى تقدير معنوية معالم النموذج اللوجستي المتعدد فحصل على النتائج التالية:

جدول (9–7): نتائج تقدير معالم الانحدار اللوجستي (أمثال النموذج اللوجستي وخواصها):

المقدرات المتحولات	β	S. E	Z Wald	df	Sig	Exp (B)	Lower of Ic 0.95	Upper of Ic 0.95
Constant	0.912	0.625	2.133	1	0.144	2.490	–	–
X_1 – جنس المريض	−4.206	1.071	15.426	1	0.000	0.015	0.002	0.122
X_2 – عمر المريض	−0.227	0.430	0.278	1	0.598	0.797	0.343	1.853
X_3 – وزن المريض	−1.264	0.435	8.461	1	0.004	0.283	0.121	0.662
X_4 – نسبة الكريات الحمراء	−1.416	0.878	2.603	1	0.107	0.243	0.043	1.356
X_5 – هيموغلوبين الدم	0.679	0.852	0.635	1	0.425	1.973	0.371	10.494
X_6 – معدل الكريات البيضاء	−0.499	0.459	1.179	1	0.278	0.607	0.247	1.494
X_7– سرعة ترسب	−0.183	0.407	0.202	1	0.653	0.833	0.375	1.850
X_8 – عدد الصفائح	1.316	0.404	10620	1	0.001	3.730	1.690	8.233

وعند دراسة هذا الجدول نجد أن عناصر β في العمود الثاني, هي نفسها العناصر التي استقرت عليها الحلول في السطر الأخير من الجدول (9–6) السابق. وهذا يجعلنا نكتب النموذج على شكل العلاقة (9–45) كما يلي:

$$ln(odds) = ln\left(\frac{P_1(X)}{1 - P_1(X)}\right) = 0.912 - 4.206X_1 - 0.227X_2 - 1.264X_3 - 1.416X_4 + \\ +0.679X_5 - 0.499X_6 - 1.183X_7 - 1.316X_8$$

ومنها يمكننا حساب الاحتمال $P_1(X)$ وهو احتمال الانتماء إلى G_1 وكتابته حسب العلاقة (9–53) كما يلي:

$$P_1(X) = \frac{1}{1 + e^{-(0.912-4.206X_1-0.227X_2-\cdots\ldots\ldots+1.316X_8)}}$$

ومنها أيضاً نقوم بحساب احتمال الانتماء إلى G_0 من العلاقة:

$$P_0(X) = 1 - P_1(X) = \frac{1}{1 + e^{+(0.912-4.206X_1-0.227X_2-\cdots\ldots\ldots+1.316X_8)}}$$

ولاختبار جودة التوفيق استخدم معيار نسبة الامكانية العظمى ورمز لها بـ $\chi^2 = -2\,ln\left[\frac{L_0}{L_1}\right]$ ، وهي تتبع تقاربياً للتوزيع χ^2 وبدرجة حرية $(p_1 - p_0)$. حيث: p_0 و p_1 هما أبعاد L_0 و L_1 على الترتيب.

وحيث أن L_0 هي قيمة دالة الامكانية العظمى عند الفرضية H_0 في G_0 .

وإن L_1 هي قيمة دالة الامكانية العظمى عند الفرضية H_1 في G_1 . فنجد أن:

جدول (9–8): اختبار x^2 للتوفيق:

	$\chi^2 = -2\ ln\left[\frac{L_0}{L_1}\right]$	df	Sig
Model	65.607	8	0.008

وهذا يدل على معنوية النموذج بشكل عام (لأن قيمة sig أقل بكثير من 0.05).

وأخيراً نلاحظ أن الجدول (9–7) السابق يعطينا عموداً خاصاً باسم Wald، وهو عبارة عن مؤشر Wald لتعريف معنوية ومصداقية تقدير كل من الأمثال β_i وهو يحسب من العلاقة التالية:

$$Wald_i = \left(\frac{\tilde{\beta}_i}{S.E(\tilde{\beta}_i)}\right)^2$$

ومن خلال عمود sig نلاحظ أن هناك ثلاثة متحولات فقط ، ذات تأثير معنوي وهي X_1 وX_3 وX_8، لأن قيم sig المقابلة لها أقل من 0.05، أما بقية المتغيرات فليس لها تأثيرات ذات أهمية أو معنوية.

كما نلاحظ أن العمود الذي يضم $EXP(\beta_1) = e^{\beta_i}$ ، هو الذي يعبر عن مقدار زيادة تابع الاستجابة Y أو التابع $logit\ (P)$ عندما يزداد المتحول المرافق لـ β_i بمقدار واحد، فمثلاً نجد أن:

$$eXP(\beta_1) = e^{-4.206} = 0.015$$

وهو يعني أن تابع الاستجابة Y سيزداد بمقدار 0.015 إذا تعبر المتحول الأول X_1 بمقدار (1)، أي إذا تغير نوع الجنس من ذكر إلى أنثى.

وبناء على صيغة النموذج الأخيرة والمحددة نقوم بحساب الاحتمالات $P_1(X)$ لكل مفردة i من مفردات العينة ثم نقوم بحساب الاحتمالات المكملة $P_0(X)$ لكل مفردة i من العلاقة التالية:

$P_0(X) = 1 - P_1(X)$، ثم نقوم بمقارنة هذين الاحتمالين لكل مفردة i . ونصنف كل مفردة i كما يلي:

القاعدة: فإذا كانت $P_{1i}(X) \geq P_{0i}(X)$ فإننا ننسب تلك المفردة i إلى G_1 .

أما إذا كانت $P_{1i}(X) < P_{0i}(X)$ فإننا ننسب تلك المفردة i إلى G_0 .

وذلك كما فعلنا على الشكل (9–2) السابق.

وبعد إجراء كل هذه العمليات وتصنيف كل مفردات العينة في إحدى المجموعتين G_1 أوG_0، نقوم من جديد بتبويب النتائج الجديدة للتصنيف ومقارنتها مع نتائج التصنيف الأصلي المستخدم في الإدارة . وعند إجراء ذلك التبويب حاسوبياً حصل الباحث على الجدول التالي:

جدول (9–9): نتائج التصنيف ونسبها المئوية

مجموع الأعداد	النسبة المئوية	التصنيف المستنبط من النموذج			
n'_i	%	G_1	G_0		

التصنيف	G_0	57	23	71.3	80
الأصلي	G_1	21	59	73.8	80
الإداري	−	49.75	51.25	72.5	−
مجموع الأعداد	n_i	78	82	−	160

ومن هذا الجدول نستنتج أن احتمال التصنيف الصحيح في المجموعة G_0 فقط كان يساوي 71.3% , وفي المجموعة G_1 كان يساوي 73.8% , ولكن الاحتمال الاجمالي للتصنيف الصحيح كان يساوي 72.5%, وهذا يعني أن المعدل الاجمالي للتصنيف الخاطئ يساوي 27.5% . ولتقدير جودة التصنيف نحسب المؤشر kappa فنجد أن:

$$kappa = \frac{n * (\sum n_{ii}) - \sum n_i n_i'}{n^2 - \sum n_i n_i'} = \frac{160(57+59) - [(80*78)+(80*82)]}{(160)^2 - [(80*78)+(80*82)]} = 0.45$$

وهي قيمة ضعيفة نسبياً، وتدل على جودة ضعيفة لعملية التصنيف المجراة في ذلك البحث.

9-7: إضافات رياضية عن الأرجحية (odds):

لتوضيح مفهوم الأرجحية وعلاقتها بالاحتمالات للظواهر الثنائية، نفترض إننا نريد معرفة معدلات الإصابة بإحدى الأمراض (كالسكري مثلاً) بين الأشخاص المعرضين له (وراثياً وصحياً وسلوكياً)، فأخذنا عينة عشوائية مؤلفة من (1000) شخص من مجتمع المعرضين لذلك المرض، وأجرينا عليهم الفحوصات المخبرية والسريرية، ثم قمنا بتبويب نتائج هذه الاختبارات حسب حالة المريض الفعلية (D^+= مصاب فعلاً وD^-= غير مصاب)، وحسب نتيجة الاختبار الإيجابية (T^+= مصاب مخبرياً) والسلبية (T^-= غير مصاب مخبرياً)، وضعناها في الجدول التالي:

جدول (9-10): نتائج تبويب المرضى حسب حالة المريض الفعلية ونتيجة الاختبار (فرضية)

حالة المريض	حالة المريض الفعلية		المجموع
نتيجة الاختبار	D^+= مصاب بالمرض	D^-= غير مصاب	
T^+ = الإصابة إيجابية	$a = 200$	$b = 85$	$n_1' = 285$
T^- = عدم إصابة	$c = 15$	700d =	$n_2' = 715$
المجموع	$n_1 = 215$	$n_2 = 785$	$n = 1000$

واعتماداً على الجدول السابق (6-10) نلاحظ أن حجم العينة $n = 1000$ شخص وهي تتألف من مجموعتين: مجموعة المصابين فعلاً (D^+) وحجمها $n_1 = 215$ شخصاً، ومجموعة غير المصابين (D^-) وحجمها $n_2 = 785$ شخصاً.

ولكن نتائج الاختبارات أظهرت لنا أن عدد المصابين مخبرياً T^+ يساوي ($n_1' = 285$) شخصاً، وعدد غير المصابين مخبرياً $n_2' = 715$ شخصاً.

وبناءً على ذلك يمكننا أن نعرف الاحتمالات التالية (وهي تحسب من هوامش الجدول):

$$P(D^+) = \frac{n_1}{n} = \frac{215}{1000} = 0.215 = \tilde{P}$$ احتمال أن يكون الشخص مصاباً :

$$P(D^-) = \frac{n_2}{n} = \frac{785}{1000} = 0.785 = \tilde{q}$$ احتمال أن يكون الشخص غير مصاب :

$$P(T^+) = \frac{n'_1}{n} = \frac{285}{1000} = 0.285$$ احتمال أن تكون نتيجة الاختبار إيجابية :

$$P(T^-) = \frac{n'_2}{n} = \frac{715}{1000} = 0.715$$ احتمال أن تكون نتيجة الاختبار سلبية :

(74 − 9)

وهنا نلاحظ أن: $P(D^+) + P(D^-) = 1$ وأن: $P(T^+) + P(T^-) = 1$

وبناء على ذلك يمكننا أن نعرف الأرجحيات (odds) المختلفة. لذلك نضع التعريف العام لأرجحية تحقق حادث (A) خل n تجربة عليه كما يلي:

$$\text{قيمة الأرجحية للحادث}(A) = \frac{\text{عدد مرات تحقق الحادث}(A)\text{ خلال التجربة}}{\text{عدد مرات عدم تحقق الحادث}(A)\text{ خلال التجربة}} = \frac{n_1(A)}{n_2(\bar{A})}$$

ونرمز لها بالرمز odds (A) ونكتبها رياضياً كما يلي:

$$odds(A) = \frac{n_1(A)}{n_2(\bar{A})} \quad : \quad \left[\text{ونكتب ذلك كمايلي: } n_1(A) : n_2(\bar{A})\right] \qquad (75-9)$$

حيث أن: $\boldsymbol{n_1(A) + n_2(\bar{A}) = n}$

وبناء على ذلك يمكننا أن نحسب قيم الأرجحيات المختلفة لحالات المرضى ولنتائج الاختبارات من بيانات الجدول (9–10) كما يلي:

$$odds(D^+) = \frac{n_1}{n_2} = \frac{215}{785} = \frac{43}{157}$$ يوجد **43** مريضاً مقابل كل **157** غير مريض :

$$odds(D^-) = \frac{n_2}{n_1} = \frac{785}{215} = \frac{157}{43}$$ يوجد **157** غير مريض مقابل كل **43** مريضاً :

$$odds(T^+) = \frac{n'_1}{n'_2} = \frac{285}{715} = \frac{57}{143}$$ هناك **57** نتيجة إيجابية مقابل كل **143** نتيجة سلبية :

$$odds(T^-) = \frac{n'_2}{n'_1} = \frac{715}{285} = \frac{143}{57}$$ هناك **143** نتيجة سلبية مقابل كل **57** نتيجة إيجابية :

(76 − 9)

كما يمكننا ببساطة استخراج العلاقات التي تربط هذه الأرجحيات بالاحتمالات (6–79) السابقة، وذلك بتقسيم بسط ومقام كل أرجحية على حجم العينة (n) فنجد أن:

$$odds(D^+) = \frac{\frac{n_1}{n}}{\frac{n_2}{n}} = \frac{P(D^+)}{P(D^-)} = \frac{p}{q} = \frac{p}{1-p} = \frac{0.215}{0.785} = 0.274$$

أي أنه يوجد مقابل كل 274 مصاب بالمرض 1000 شخص غير مصاب به.

$$odds(D^-) = \frac{\frac{n_2}{n}}{\frac{n_1}{n}} = \frac{P(D^-)}{P(D^+)} = \frac{q}{p} = \frac{1-p}{p} = \frac{1}{odds(D^+)} = 3.65 \qquad (77-9)$$

أي أنه يوجد مقابل كل 365 غير مصاب يوجد 100 مصاب.

$$odds(T^+) = \frac{\frac{n'_1}{n}}{\frac{n'_2}{n}} = \frac{P(T^+)}{P(T^-)} = \frac{p'}{q'} = \frac{p'}{1-p'} = \frac{0.285}{0.715} = 0.363$$

أي أنه مقابل كل 363 نتيجة إيجابية يوجد 1000 نتيجة سلبية.

$$odds(T^-) = \frac{n'_1/n}{n'_2/n} = \frac{P(T^-)}{P(T^+)} = \frac{q}{p'} = \frac{1-p'}{p'} = \frac{1}{odds(T^+)} = 2.75$$

أي أنه مقابل كل 275 نتيجة سلبية يوجد 100 نتيجة إيجابية.

ومما سبق يمكننا استنتاج أن أرجحية أي حادث (A) ترتبط باحتمال تحققه وعدم تحققه وفق العلاقة التالية:

$$odds(A) = \frac{P(A)}{P(\bar{A})} = \frac{P(A)}{1-P(A)} \qquad (\mathbf{78-9})$$

وأن نستنتج أن: $\boldsymbol{odds(\bar{A}) = \frac{1}{odds\,(A)}}$ **وأن:** $\boldsymbol{odds(\bar{A}) = \frac{P(\bar{A})}{P(A)}}$

$$odds(A) * odds(\bar{A}) = 1 \qquad (79-9)$$

كما يمكننا أن نعرّف عدداً من المؤشرات الاحصائية اعتماداً على التكرارات الداخلية والتكرارات الهامشية والمبينة في الجدول (9–10) السابق، وهذه المؤشرات هي الاحتمالات الشرطية التالية:

1- الحساسية (sensibility): وهي نسبة عدد المرضى (D^+) الذين اعتبرتهم الاختبارات إنهم مصابين بالمرض (أي كانت نتيجة الاختبار T^+ متطابقة مع الحالة الفعلية للمريض D^+) ونرمز لها بالرمز $P(T^+/D^+)$ ونحسبها من الاحتمال الشرطي التالي:

$$P(T^+/D^+) = \frac{a}{a+c} = \frac{a}{n_1} = \frac{200}{215} = 0.93023 \qquad (80-9)$$

وهو احتمال أن تكون نتيجة اختبار المريض متطابقة مع حالته المرضية (أي أن تكون نتيجة الاختبارات الإيجابية صحيحة).

2- الخصوصية أو النوعية (specificity): وهي نسبة عدد غير المرضى (D^-) الذين اعتبرتهم الاختبارات إنهم غير مصابين (أي كانت نتيجة الاختبارات السلبية T^- متطابقة مع الحالة الفعلية للمريض D^-)، ونرمز لها بـ $P(T^-/D^-)$ ونحسبه من الاحتمال الشرطي التالي:

$$P(T^-/D^-) = \frac{d}{b+d} = \frac{d}{n_2} = \frac{700}{785} = 0.89172 \qquad (81-9)$$

وهو احتمال أن تكون نتيجة اختبار غير المريض متطابقة مع حالته الصحية (أي أن تكون نتيجة الاختبارات السلبية الصحيحة).

3- نسبة الاختبارات الإيجابية الكاذبة (غير الصحيحة): وهي نسبة عدد غير المرضى (D^-) الذين اعتبرتهم الاختبارات مصابين بالمرض (T^+) ونرمز لها بالرمز $P(T^+/D^-)$ ونحسبها من الاحتمال الشرطي التالي:

$$P(T^+/D^-) = \frac{b}{b+d} = \frac{85}{285} = 0.10828 \qquad (82-9)$$

وهو احتمال أن تكون نتيجة اختبار غير المريض إيجابية (النتيجة غير صحيحة).

4- نسبة الاختبارات السلبية الكاذبة (غير الصحيحة): وهي نسبة عدد غير المرضى (D^+) الذين اعتبرتهم الاختبارات أنهم غير مصابين (T^-) ونرمز لها بالرمز $P(T^-/D^+)$ وتحسب من الاحتمال الشرطي التالي:

$$P(T^-/D^+) = \frac{c}{a+c} = \frac{15}{215} = 0.06977 \qquad (83-9)$$

وهو احتمال أن تكون نتيجة اختبار المريض سلبية (النتيجة غير صحيحة).

5- مصداقية الاختبارات: وهي نسبة كل الاختبارات الصحيحة (الإيجابية والسلبية) إلى المجموع الكلي للاختبارات (n) ويُرمز لها بـ AT وتحسب من العلاقة:

$$AT = \frac{a+d}{n} = \frac{a+d}{a+b+c+d} = \frac{200+700}{1000} = 0.90 \qquad (84-9)$$

6- معدل التصنيف الخاطئ: وهو نسبة كل الاختبارات غير الصحيحة إلى المجموع الكلي للاختبارات (n) ويرمز له بـ MR وتحسب من العلاقة:

$$MR = \frac{b+c}{n} = \frac{b+c}{a+b+c+d} = \frac{85+15}{1000} = 0.10 \qquad (85-8)$$

وهنا نلاحظ أن: $MR = 1 - AT$

كما يمكننا تعريف عدداً من الأرجحيات الشرطية، وذلك بناءً على التكرارات الداخلية فقط المبينة في الجدول (9–10) السابق وهذه الأرجحيات هي:

$$odds(T^+/D^+) = \frac{a}{c} = \frac{200}{15} = \frac{40}{3} \qquad (86-9)$$

أي أنه يوجد 40 نتيجة إيجابية مقابل كل 3 نتائج سلبية وذلك عند مجموعة المرضى D^+ .

$$odds(T^+/D^-) = \frac{b}{d} = \frac{85}{700} = \frac{17}{140}$$

أي أنه يوجد 17 نتيجة إيجابية مقابل كل 140 نتيجة سلبية عند مجموعة غير المرضى D^- .

$$odds(T^-/D^+) = \frac{c}{a} = \frac{15}{200} = \frac{3}{40}$$

أي يوجد 3 نتائج سلبية مقابل كل 40 نتيجة إيجابية عند مجموعة المرضى D^+ .

$$odds(T^-/D^-) = \frac{d}{b} = \frac{700}{85} = \frac{140}{7}$$

يوجد 140 سلبية مقابل كل 7 إيجابية عند غير المرضى

$$odds(D^+/T^+) = \frac{a}{b} = \frac{200}{85} = \frac{40}{17}$$

يوجد 40 مريض مقابل كل 17 في التحاليل الإيجابية

$$odds(D^+/T^-) = \frac{c}{d} = \frac{15}{700} = \frac{3}{140}$$

يوجد 3 مريض مقابل كل 140 في التحاليل السلبية

يوجد 17 غير مريض مقابل كل 40 في التحاليل الإيجابية $$odds(D^-/T^+) = \frac{b}{a} = \frac{85}{200} = \frac{17}{40}$$

يوجد 140 غير مريض مقابل كل 3 في التحاليل السلبية $$odds(D^-/T^-) = \frac{d}{c} = \frac{700}{15} = \frac{140}{3}$$

وأخيراً نعرف المؤشرات الهامة التالية:

7- نسبة الأرجحية (odds Ratio): وتعرف للمرضى ككل ونرمز لها بـ OR وتحسب من العلاقة التالية:

$$OR = \frac{odds(D^+/T^+)}{odds(D^+/T^-)} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{a*d}{b*c} = \frac{200*700}{85*15} = 109.80 \qquad (87-9)$$

أما نسبة الأرجحية لغير المرضى فتحسب من العلاقة:

$$\overline{OR} = \frac{odds(D^-/T^+)}{odds(D^-/T^-)} = \frac{\frac{b}{a}}{\frac{d}{c}} = \frac{b*c}{a*d} = \frac{1}{OR} = 0{,}009107 \qquad (88-9)$$

8- نسبة المخاطرة (Risk Ratio): وهي نسبة احتمال الحادث المطلوب على احتمال الحادث المتمم له ونرمز لها بـ RR وتحسب لاختبارات المرضى من العلاقة:

$$RR = \frac{P(D^+/T^+)}{P(D^+/T^-)} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{ac+ad}{ac+bc} = \frac{200*15+200*700}{200*15+85*15} = 33.45 \qquad (89-9)$$

وعندما يكون الحد $(a*c)$ صغيراً بالنسبة لـ $(a*d)$ فيمكن اهماله وتصبح نسبة المخاطرة المقربة كما يلي:

$$\widetilde{RR} \approx \frac{ad}{bc} \approx OR \qquad (90-9)$$

9- نسبة المخاطرة المطلقة لاختبارات المرضى وتحسب من العلاقة:

$$AR = P(D^+/T^+) - P(D^+/T^-) = \left[\frac{a}{a+b} - \frac{c}{c+d}\right] = 0.68078 \qquad (91-9)$$

الفصل العاشر
الاختبارات اللامعلمية

10-1 : تمهيد:

لقد لاحظنا من خلال الفصول السابقة أن اختبارات الفرضيات هي مؤشرات حساسة لكشف صحة أو عدم صحة فرضية العدم H_0. ولكن تلك الاختبارات لا تتأثر فقط بوضع الفرضية H_0 نفسها، بل تتأثر بتحقق أو عدم تحقق الشروط المرافقة للاختبار المفروض، وكلما كانت تلك الشروط محققة، كان الاختبار أكثر فاعليةً وقوةً. ولقد رأينا سابقاً أن الاختبارات التي استخدمناها في التحقق من الفرضيات الموضوعة حول متوسط أو إجمالي المجتمع أو حول النسبة فيه ..إلخ، كانت تشترط أن تكون الخاصة المدروسة في المجتمع خاضعة للتوزيع الطبيعي، كما تشترط أن تكون العينات عشوائية، أو أن تكون المجتمعات متجانسة، أو أن تكون الخاصة قابلة للقياس ..إلخ. ولذلك يجب التأكد من تحقق مثل هذه الشروط (أو مثيلاتها) في البيانات المدروسة قبل تطبيق مؤشر الاختبار عليها واستخلاص النتائج الممكنة منها.

وفي حالة عدم تحقق مثل هذه الشروط أو بعضها، فإن الاختبار يفقد كثيراً من الثقة فيه ويصعب الاعتماد على نتائجه. وبعبارةٍ أخرى، فإن تطبيق الاختبارات المعلمية على الحالات التي لا تكون فيها الشروط محققة يعتبر أمراً خطيراً للغاية، وذلك ليس بفعل البيانات الإحصائية، بل بسبب عدم تحقق أحد أو بعض الشروط المرافقة للاختبار.

وإضافةً إلى عدم تحقق بعض الشروط السابقة، يواجه الإحصائي في عمله الكثير من المتحولات غير القابلة للقياس الكمّي , مثل متحولات : النوع والجنس، ودرجة التعليم، ومكان السكن والولادة، والحالة الزواجية، والحالة التعليمية ...إلخ. وتسمى هذه المتحولات بالمتحولات النوعية , ونحتاج عند التعامل معها إلى أساليب جديدة تتناسب مع تلك المتحولات، وتتفق مع نوع المعلومات المتوفرة عنها. ولذلك تم تصنيف المتحولات العشوائية إلى: كمية ونوعية، ونعرفهما بما يلي:

1 . المتحولات الكمّية:

وهي المتحولات التي يمكن وصفها بواسطة القياس الكمّي، والتعبير عن خواصها بواسطة أعداد مرفقة بواحدات قياس معينة. وتقسم هذه المتحولات إلى صنفين أساسيين هما:

. متحولات منقطعة: وهي المتحولات التي تأخذ قيم منقطعة صحيحة، أو غير صحيحة، مثل: عدد أفراد الأسرة، عدد الطلاب، عدد السيارات...إلخ.

. متحولات مستمرة: وهي المتحولات التي تأخذ قيماً مستمرة ومتلاصقة، ويمكنها أن تأخذ أية قيمة تقع بين أي قيمتين متجاورتين مهما كانتا قريبتين، مثل: درجة الحرارة، العمر، الطول ...إلخ.

ويتم قياس المتحولات الكمّية بواسطة سلمين هما:

. السلم المجالي Interval scale: وهو سلم كمّي منسوب إلى نقطة صفر افتراضية لتحديد بداية القياسات، وتكون الأبعاد بين البيانات مقاسة بوحدة قياس محددة لكل متحول (كالمسافة ودرجة الحرارة).

. السلم النسبي Rational scale: وهو سلم كمّي منسوب إلى نقطة صفر حقيقية هي المبدأ الحقيقي للقياسات وتنسب إليها جميع القياسات، مثل الوزن، والطول..إلخ، المنسوبة إلى نقطة الصفر المطلق(عدم وجود وزن أو طول). وإن النسبة في السلمين بين أي قيمتين متشابهتين تكون مستقلة عن وحدة القياس.

2 . المتحولات النوعية:

وهي المتحولات التي لا يمكن وصفها إلا بواسطة حالات معنية مختلفة بعضها عن بعض بصفة معينة، أو بدرجة محددة، مثل: الجنس، العرق، درجة التعليم، الحالة الزواجية، الحالية العملية، الحالة الصحية ...إلخ. وتم تصنيف المتحولات النوعية حسب سلم القياس فيها إلى مستويين هما:

. السلم الاسمي Nominal scale: وهو سلم نوعي ومن أضعف مستويات القياس المستخدمة في الإحصاء، وفيه يتم تصنيف الحالات التي يأخذها المتحول ضمن فئات، أو مجموعات اسمية، وترفق بالتكرارات المطلقة أو النسبية المقابلة لكل منها، وتوضع في جدول مناسب.

وخير مثال على المتحولات التي تخضع لهذا السلم هي المتحولات التالية:

الجنس: وله حالتان: ذكر، أنثى.

العِرق: وله عدة حالات: الأبيض، الأصفر، الأسود , الاحمر.

الحالة الزواجية: ولها عدة حالات: عازب، متزوج، مطلق، أرمل.

ويشترط هنا أن تكون جميع العناصر ضمن كل حالة متكافئة فيما بينها، وأن لا تكون أفضلية لحالة على أخرى، وهذا يعني أن جميع الذكور ضمن فئة الذكر متساوون، وأن جميع الإناث ضمن فئة الأنثى متكافئات، وأنه لا يمكننا أن نفضل فئة الذكور على فئة الإناث، أو بالعكس...إلخ.

ويمكن الاعتماد على هذا السلم عند تطبيق اختبارات حول النسب المئوية وحول التوافق والتجانس بين الحالات لمتحولين.

. السلم الرتبي Ordinal scale: وهو سلم نوعي شبيه بالسلم الاسمي، وفيه يتم تصنيف الحالات التي يأخذها المتحول ضمن فئات اسمية محددة، ولكنه يمكن تفضيل أو تمييز الحالات المختلفة التي يأخذها المتحول والقيام بترتيبها حسب علاقة ترتيب معينة، مثل: أكبر، أو أصغر، أو أفضل ...إلخ. ويندرج تحت هذا السلم المتحولات التالية:

درجة التعليم وحالاتها: أمي، ابتدائي، إعدادي، ثانوي، جامعي، عليا.

الرتبة الوظيفية وحالاتها: ممتازة، أولى، ثانية، ثالثة، رابعة، خامسة.

وهنا نشير إلى أن العمليتين المعرفتين على هذا السلم هما: عملية التكافؤ ضمن الحالات، والترتيب بين الحالات.

ولهذه الأسباب تم تصنيف اختبارات الفرضيات إلى صنفين أساسيين هما:

. اختبـارات معلميـة Parametric statistics: وهـي الاختبـارات التـي تتنـاول المتحـولات الكمّيـة فـي المجتمـع المـدروس وتختبـر أهم معالمـه. مثـل اختبـارات المتوسـط والإجمـالي والنسـبة والتبـاين والفـرق بـين المتوسطين، والفرق بين نسبتين ....إلخ.

وأشهر هذه الاختبارات هي الاختبارات التي تناولناها في الفصول السابقة, مثل اختبارات z و t بجميع أشكالهما.

. اختبارات لا معلمية Non-parametric statistics: وهي الاختبارات التي تتناول المتحولات النوعية فـي المجتمـع المـدروس. وهـذه الاختبـارات لا تعتمـد علـى القياسـات بـل علـى التكرارات المقابلـة للحـالات أوعلى الرتب الممكنة لتلك المتحولات. وإن أشهر هذه الاختبارات هو اختبار χ^2 .

وحتى تكتمل موضوعات الكتاب سنخصص هذا الفصل لاستعراض بعض الاختبارات اللامعلمية ، التي تتعامل مع التكرارات المقابلة لكل حالة من الحالات النوعية أو مع الرتب الموافقة لكل درجة من الدرجات المختلفة للظاهرة المدروسة.

ومن أهم مميزات الاختبارات اللامعلمية أنها لا تشترط توفر شروط افتراضية كثيرة في المجتمع أو العينـة. وإن أغلبها يكتفي بتوفر شروط ذات طبيعة بديهية مثل:

. أن تكون العينة عشوائية، أو أن تكون التجارب مستقلة.

. أن تكون العناصر متكافئة ضمن كل حالة، ولا يمكن تفضيل حالة عن أخرى.

. أن تكون الحالات ضمن كل متحول منفصلة(غير متقاطعة) ومتكاملة.

كما أن هذه الاختبارات تتميز بالبساطة وبسهولة وسرعة الحسابات، ولا يحتاج الباحثون عند تطبيقها إلى أساس رياضي كبير لفهمها، وهي تعطي نتائج سريعة ومضمونه، وخاصة عندما تكون شروط الاختبارات المعلمية غير محققة.

إن مجـالات تطبيق الاختبـارات اللامعلميـة لا تنحصـر فـي التعامـل مـع المتحـولات والبيانـات النوعيـة، بـل يمكن تطبيقها أيضاً على البيانات الكمية، حيث يمكننا اعتبار القيم أو المجالات حالات نوعية، إلا أنه لاينصح باستبدال الاختبارات المعلمية بالاختبارات اللامعلمية وتطبيقها على المتحولات والبيانات الكمية، إلا إذا كانت الشروط المرافقة للاختبار المعلمي غير محققة.

ولقد شهدت الحقبة الأخيرة تطوراً كبيراً للاختبارات اللامعلمية، فظهر العديد من المؤشرات الاختبارية التي تعالج مختلف الحالات العلمية والاقتصادية والاجتماعية... إلخ. لابل تعددت الأشكال والصيغ التي تعالج نفس المشكلة أو تطبق على بيانات نوعية متشابهة. ومنها ما نجح فانتشر, ومنها ما فشل فاندثر.

ولكن تعدد تلك المؤشرات بأشكالها وأنواعها وضع الباحثين أمام مشكلتين جديدتين هما:

. مشكلة تحديد الاختبار اللامعلمي المناسب لتطبيقه على البيانات المتوفرة .

. مشكلة التفضيل بين الاختبارات اللامعلمية التي تنطبق على الحالة نفسها.

ولمعالجة هاتين المشكلتين تم وضع بعض المعايير المحددة للاعتماد عليها عند التحديد والتفضيل، وأهم هذه المعايير هي*:

. مقاييس البيانات Measures of data.

. نوع عينة البيانات وحجمها Type and size of sample.

. شكل البيانات والهدف من الاختبار Form of data.

. قوة الاختبار $(1-\gamma)$: Power of test.

.الفعالية أو الكفاءة النسبية Relative efficiency.

. عدم تحيز الاختبار Unbiasedness of test.

. تماسك الاختبار Consistency of test.

. تحفظ الاختبار Conservation of test.

وتعتبر قوة الاختبار والتي تساوي $(1-\gamma)$ أهم هذه المعايير، ولكن عملية حسابها تحتاج إلى مقدمات وبراهين متعمقة لا تدخل ضمن منهاج هذا الكتاب ** ، ولهذا فإننا سنستعرض أهم الاختبارات اللامعلمية وفق الترتيب الآتي:

10-2: الاختبارات اللامعلمية المعرفة على المتحولات النوعية الأسمية (من عينة واحدة):

عندما نسحب عينة عشوائية واحدة لإجراء دراسة إحصائية معينة، فإن استمارة البحث تكون متضمنة عدداً محدداً من الأسئلة عن مؤشرات كمية وعدداً آخر عن مؤشرات نوعية، واعتماداً على هذه العينة يمكننا أن نقوم بتصنيف الأجوبة ضمن جداول مختلفة تتضمن البيانات والتكرارات المقابلة لها، ثم ننتقل إلى دراسة خصائص المؤشرات السابقة، وحساب متانة العلاقات الارتباطية بينها، وسنستعرض فيما يلي هذه القضايا حسب أنواع المؤشرات وفق التالي:

وهنا يمكن أن نميز عدة أنواع من المتحولات الأسمية هي:

. متحولات أسمية ثنائية: ويكون لها حالتان فقط مثل: الجنس والحالة الصحية.

. متحولات أسمية متعددة: ويكون لها عدة حالات مثل: الدرجة التعليمية والحالة العملية.

وإن الاختبارات التي سندرسها هنا هي:

10-2-1: اختبارات الاقتران والتوافق بين متحولين اسميين ثنائيين Association Tests:

وهي اختبارات لامعلمية معرفة على متحولين اسميين x و y، لكل منهما حالتان فقط. ولهما التكرارات المقابلة كما في الجدول الآتي:

---

* انظر: *Segiel P.18-31, Conover P.83-90*

** انظر: العلي، ابراهيم، الإحصاء الرياضي، جامعة حلب، 1986 ص.153.

جدول(10−1): جدول التكرارات الرباعي

y \ x	x_1	x_2
y_1	A	B
y_2	C	D

ولإجراء بعض الاختبارات على المتحولين **x , y** اللذين في هذا الجدول نضع الفرضيتين كما يلي:

فرضية العدم H_0 : لايوجد اقتران بين المتحولين, الفرضية البديلة: يوجد اقتران بين المتحولين.

ويعرف على المتحولين, اللذين في هذا الجدول عدة اختبارات لامعلمية أهمها:

1: معامل الاقتران الرباعي:

يستخدم هذا المعامل لاختبار وجود اقتران بين هذين المتحولين بطريقة حساب معامل الاقتران Association coefficient، المعرف بالعلاقة التالية:

$$CA = \frac{A \cdot D - B \cdot C}{A \cdot D + B \cdot C} \qquad (1-10)$$

وهو يأخذ قيماً محصورة في المجال $]-1\ ,\ +1[$، وكلما كانت قيمته المطلقة قريبة من الواحد، كان الاقتران بين المتحولين قوياً. أما عندما تكون قيمته المطلقة قريبة من الصفر، أو أصغر من 0,50، فلا نأخذ بالاقتران بين المتحولين.

وحتى لا يحصل أي التباس حول ذلك, يفضل عند إجراء الدراسات الدقيقة القيام باختبار معنوية هذا المعامل من خلال مؤشر الاختبار الطبيعي الآتي:

$$z = \frac{2(CA)}{1-(CA)^2} \cdot \frac{1}{\sqrt{\frac{1}{A}+\frac{1}{B}+\frac{1}{C}+\frac{1}{D}}} \qquad (2\text{-}10)$$

وبمقارنة قيمة Z مع قيمة Z الجدولية المقابلة لمستوى الدلالة المفروض، نقرر فيما إذا كانت قيمة معامل الاقتران معنوية، أم لا. وبالتالي نقبل فرضية العدم أم نرفضها.

مثال(10−1):

ادرس فيما إذا كان هناك اقتران بين الجنس والنجاح من خلال نتائج 585 طالباً في إحدى المدارس الثانوية, إذا علمت أن النتائج الامتحانية أعطتنا الأعداد التالية:

جدول(10−2): بيانات المثال

النتيجة \ الجنس	ذكر	أنثى
ناجح	**215**	**150**
راسب	**130**	**90**

الحل: نقوم بحساب معامل الاقتران الرباعي بين هذين المتحولين الاسميين، فنجد أن:

$$CA=\frac{A\cdot D-B\cdot C}{A\cdot D+B\cdot C}=\frac{215\cdot 90-150\cdot 130}{215\cdot 90+150\cdot 130}$$

$$CA=\frac{-150}{38850}=-0{,}00386$$

وهي قيمة صغيرة جداً تدل على عدم وجود اقتران بين الجنس والنتيجة. وللتأكد من معنوية هذه القيمة نحسب قيمة مؤشر الاختبار Z، فنجد أن:

$$Z=\frac{2(0{,}00386)}{1-(0{,}00386)^2}\cdot\frac{1}{\sqrt{\frac{1}{215}+\frac{1}{150}+\frac{1}{130}+\frac{1}{90}}}=0{,}0772\cdot\frac{1}{\sqrt{0{,}030}}$$

$$Z=0{,}446$$

وبمقارنة هذه القيمة مع القيمة الجدولية لـ Z عند مستوى دلالة $0{,}05$، والتي تساوي $z_{1-\frac{\alpha}{2}}=1{,}96$ ، نجد أن $Z<Z_{1-\frac{\alpha}{2}}$، لذلك نقبل فرضية العدم H_0 حول CA والتي تقول إن قيمة هذا المعامل معدومة ولا يوجد أي اقتران بين متحولي الجنس ونتيجة الامتحان.

2: معامل التوافق الرباعي (2×2):

يطبق هذا المعامل على متحولين اسميين ثنائيين (أو كميين ثنائيين)، ولهما التكرارات المقابلة كما في الجدول الرباعي الآتي:

جدول(10−3): تكرارات الجدول الرباعي

y \ x	x_1	x_2	المجموع
y_1	A	B	$A+B$
y_2	C	D	$C+D$
المجموع	$A+C$	$B+D$	N

ويعرف معامل التوافق الرباعي بالعلاقة التالية:

$$T=\frac{N\left(A\cdot D-B\cdot C\right)^2}{(A+B)(C+D)(A+C)(B+D)} \qquad (3-10)$$

وهو متحول خاضع لتوزيع $\chi^2(x)$ بدرجة حرية واحدة $(\nu=1)$.

ولاتخاذ قرار حول فرضية العدم (عدم التوافق) نقارن قيمة هذا المعامل مع قيمة متحول التوزيع $\chi^2(x)$ المقابلة لمستوى الدلالة ولدرجة حرية واحدة $(\nu=1)$، ونتخذ القرار المناسب كالعادة.

مثال(10−2):

لدراسة التوافق بين أجوبة عينة بحجم 380 من المبحوثين حول السؤالين التاليين:

السؤال الأول: هو هل تؤيد عمل المرأة؟ نعم / لا

السؤال الثاني: هل تؤيد تنظيم الإنجاب؟ نعم / لا

أخذنا نتائج الأجوبة وبوبناها مع تكراراتها فكانت كما في الجدول الآتي:

جدول(10−4): بيانات المثال

السؤال الأول / السؤال الثاني	نعم	لا	المجموع
نعم	160	50	210
لا	40	130	170
المجموع	200	180	380

والمطلوب دراسة التوافق بين أجوبة المبحوثين على السؤالين السابقين.

الحل:

لدراسة التوافق المطلوب نطبق العلاقة (10−3)، فنجد أن:

$$T=\frac{N\left(A\cdot D-B\cdot C\right)^2}{\left(A+B\right)\left(C+D\right)\left(A+C\right)\left(B+D\right)}=\frac{380(160\cdot 130-50\cdot 40)^2}{210\cdot 170\cdot 200\cdot 180}$$

$$T=\frac{380(353440000)^2}{1285200000}=104{,}5$$

ومن الجداول الملحقة , نجد أن قيمة متحول $\chi^2\left(x\right)$ المقابلة لمستوى الدلالة 0.05. ولدرجة حرية واحدة $\left(\nu=1\right)$، تساوي $\chi^2\left(x\right)=3{,}84$.

وبمقارنة قيمة T المحسوبة مع القيمة الحرجة 3,84، نجد أن $T>3{,}84$, لذلك نرفض فرضية العدم حول عدم التوافق ونقبل الفرضية البديلة التي تقول بوجود توافق بين أجوبة المبحوثين حول السؤالين السابقين. ونستنتج أن معظم الذين يؤيدون عمل المرأة يؤيدون أيضاً تنظيم الإنجاب، والعكس بالعكس.

3− معامل التوافق الرباعي المصحح (تصحيح ييتس):

إذا كان حجم العينة n صغيراً، وكانت المتحولات المدروسة مستمرة، فإن قيمة T المحسوبة من العلاقة السابقة (10−3) في حالة الجداول من المرتبة 2× 2، تعطينا قيمة متحيزة وأكبر من القيم الحقيقية T.

وللتخلص من ذلك التحيز قدم العالم (ييتس) تصحيحاً لذلك التحيز بإضافة $\frac{1}{2}$ إلى التكرار الذي يقل عن التكرار المتوقع، وبطرح $\frac{1}{2}$ من التكرار الذي يزيد عن التكرار المتوقع , أي كأنه قام بإضافة $\frac{1}{2}$ إلى كل

من a و d وطرح $\frac{1}{2}$ من كل من b وc، أو بالعكس. وبعد الإصلاح والمعالجة والتعويض في العلاقة السابقة حصل على العلاقة المصححة الآتية:

$$\chi_M^2 = \frac{n\left[|a\cdot d - b\cdot c| - \frac{n}{2}\right]^2}{(a+b)(c+d)(a+c)(b+d)} \qquad (4-10)$$

وعند الحساب نطرح من القيمة المطلقة للمقدار التالي $|a\cdot d - b\cdot c|$ المقدار $\frac{n}{2}$ ، نربع الناتج ونضربه بـ n ونقسمه على الجداءات المذكورة.

مثال(10-3):

لنحسب قيمة χ_M^2 المصححة لمعطيات المثال السابق فنجد أن:

$$x_M^2 = \frac{145\left[|50\cdot 25 - 40\cdot 30| - \frac{145}{2}\right]^2}{90\cdot 55\cdot 80\cdot 65} = 0{,}00285$$

وهي قيمة أصغر من القيمة الأولى لـ χ^2 قبل التعديل، كما أنها أصغر من القيمة الجدولية $\chi_\alpha^2 = 3{,}84$ ، لذلك نرفض فرضية العدم أيضاً.

4- معامل التوافق المتعدد $(k\times l)$ Contingency coefficient (من عينة واحدة)

يعتمد هذا المعامل على توزيع χ^2 الذي درسناه سابقاً, ويعرف على متحولين اسميين متعددين، أو على متحول أسمي متعدد في عدة مجتمعات أوطبقات (لكل منهما أكثر من حالتين), وبحيث تكون تكراراتهما المتقاطعة مأخوذة من عينة واحدة . ولنفترض الآن أن المتحول الأول X يأخذ k حالة، وأن المتحول الثاني Y يأخذ l حالة(أوطبقة), وأن جدول التكرارات المشتركة لحالاتهما غير المرتبة يأخذ الشكل التالي:

جدول(10-5):التكرارات المشتركة للحالات المتقاطعة

حالات المتحول الأول x غير المرتبة / حالات المتحول الثاني y غير المرتبة	x_1	x_2	x_3	…	x_j	…	x_k	المجموع
y_1	n_{11}	n_{12}	n_{13}	…	n_{1j}	…	n_{1k}	n'_1
y_2	n_{21}	n_{22}	n_{23}	…	n_{2j}	…	n_{2k}	n'_2
y_3	n_{31}	n_{32}	n_{33}	…	n_{3j}	…	n_{3k}	n'_3
y_i	n_{i1}	n_{i2}	n_{i3}	…	n_{ij}	…	n_{ik}	n'_i
y_ℓ	$n_{\ell 1}$	$n_{\ell 2}$	$n_{\ell 3}$	…	$n_{\ell j}$	…	$n_{\ell k}$	n'_ℓ
المجموع	n_1	n_2	n_3	…	n_1	…	n_k	n

ثم نقوم بحساب قيمة المتحول χ^2 المعرف بعلاقة (بيرسون) التالية:

$$\chi^2 = \sum_{i=1}^{\ell}\sum_{j=1}^{k}\frac{\left(n_{ij} - E_{ij}\right)^2}{E_{ij}} \qquad (5-10)$$

حيث أن : $n_{\ell j}$ هي التكرارات الفعلية أوالتجريبية المبينة في الجدول السابق.

وحيث أن: E_{ij} هي التكرارات المتوقعة، وتحسب من العلاقة: $E_{ij} = \frac{n_j \cdot n'_i}{n}$.

علماً بأن المتحول χ^2 يخضع للتوزيع $\chi^2(x)$ وبدرجة حرية $v = (k-1)(\ell-1)$.

ولاختبار معنوية ذلك التوافق نقارن قيمة χ^2 المحسوبة مع قيمة χ^2_α الحرجة والمقابلة لمستوى *الدلالة* a *ولدرجة الحرية* $(\ell-1)(k-1) = v$ *, ونتخذ القرار المناسب كالعادة.*

ملاحظةهامة : يمكن تطبيق هذا المعامل على متحول واحد $X:(x_1 \quad x_2 \quad x_3 \quad x_i.......x_k)$ ***في عدة مجتمعات*** أوطبقات $Y:(y_1 \quad y_2 \quad y_3 \quad y_j.......x_\ell)$ ***, وذلك بسحب عدة عينات*** **من تلك المجتمعات ودراسة توافقها حسب تغيرات تكرارات X فيها, وهذه الحالة هي التي دعت إلى تسميته بمعامل التوافق.**

ونظراً لعدم حساسية هذا المعامل, قام بعض العلماء بتطوير صيغته الرياضية السابقة للمعامل χ^2، لتعريف معاملات جديدة لدراسة التوافق بين المتحولات أو المجتمعات وسميت بأسمائهم، وأهم هذه المعاملات هي :

*- **معامل التوافق C** : ولقد استخرجه العالم (كارل بيرسون) وعرفه بالعلاقة:

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

وكلما كانت قيمة C قريبة من الواحد كان التوافق جيداً، وكان الارتباط متيناً، ولكن قيمته تبقى أقل من (1).

*- **معامل التوافق** T^2:

ولقد استنبطه العالم (تشوبيرو) اعتماداً على معامل التوافق C وتوصل إلى العلاقة التالية:

$$T^2 = \frac{C^2}{\left(1 - C^2\right)\sqrt{(k-1)(\ell-1)}}$$

حيث k عدد الأعمدة و ℓ عدد الأسطر في جدول التوافق لحساب χ^2.

*- **معامل (كرامر Cramer) للتوافق** ويعرف بالعلاقة:

$$C_1 = \sqrt{\frac{\chi^2}{n\,(q-1)}} \qquad (6-10)$$

حيث q أصغر العددين ℓ، أو k. و n عدد التكرارات.

***– معامل (ييل Yulle) للتوافق ويعرف بالعلاقة:**

$$C_2 = \frac{\chi^2}{n} \qquad (7-10)$$

وإن جميع هذه المعاملات تعطينا تقديرات لمقدار التوافق بين المتحولين الاسميين إذا كانا متعددي الحالات، وهي تأخذ قيمتها في المجال $]0 , 1[$ ، ولكنها لاتبلغ الواحد.

وكلما كانت قيمة المعامل المحسوبة قريبة من الواحد، كان التوافق جيداً، وبالعكس، كلما كانت قيمته قريبة من الصفر كان التوافق بين حالات المتحولين ضعيفاً.

مثال(10–4):

احسب معامل التوافق χ^2 ثم المعامل C حسب علاقة بيرسون بين متحولي الحالة الاجتماعية والجنس بين موظفي الجامعة، وذلك من البيانات التالية المأخوذة من عينة الدراسة ذات الحجم $n = 192$ موظفا.

جدول(10–6): التكرارت المقابلة للحالات المتقاطعة

الحالة الاجتماعية / الجنس	أعزب	متزوج	مطلق	أرمل	المجموع
ذكر	40	50	10	5	105
أنثى	30	35	15	7	87
المجموع	70	85	25	12	192

الحل: حتى نستطيع حساب قيمة χ^2 ، علينا أولاً أن نحسب التكرارات المتوقعة (النظرية) من العلاقة: $E_{ij} = \frac{n_j \cdot n'_i}{n}$ ، فنجد أن:

$$E_{11} = \frac{105 \cdot 70}{192} = 38{,}28 \quad , \quad E_{12} = \frac{105 \cdot 85}{192} = 46{,}48$$

$$E_{13} = \frac{105 \cdot 25}{192} = 13{,}67 \quad , \quad E_{14} = \frac{105 \cdot 12}{192} = 6{,}56$$

$$E_{21} = \frac{87 \cdot 70}{192} = 31{,}72 \quad , \quad E_{22} = \frac{87 \cdot 85}{192} = 38{,}52$$

$$E_{23} = \frac{87 \cdot 25}{192} = 11{,}33 \quad , \quad E_{24} = \frac{87 \cdot 12}{192} = 5{,}44$$

ومنها نقوم بحساب قيمة χ^2 من العلاقة(10–5)، فنجد أنها تساوي:

$$\chi^2 = \frac{(40-38{,}28)^2}{38{,}28} + \frac{(50-46{,}48)^2}{46{,}48} + \frac{(10-13{,}67)^2}{13{,}67} + \frac{(5-6{,}56)^2}{6{,}56}$$

$$+ \frac{(30-31{,}72)^2}{31{,}72} + \frac{(35-38{,}52)^2}{38{,}52} + \frac{(15-11{,}33)^2}{11{,}33} + \frac{(7-5{,}44)^2}{5{,}44} =$$

$$\chi^2 = 0{,}0773 + 0{,}2666 + 0{,}9853 + 0{,}3710 + 0{,}0933$$

$$+ 0{,}3217 + 1{,}1888 + 0{,}2868 = 3{,}59$$

ولاختبار معنوية معامل هذا التوافق نقارن قيمة χ^2 المحسوبة مع القيمة الحرجة χ^2_α, المقابلة لمستوى الدلالة 0.05 ولدرجة حرية $v=1*3=3$ ، والتي تؤخذ من الجداول الإحصائية وتساوي $\chi^2_\alpha = 7.81$.

وبالمقارنة نجد أن القيمة المحسوبة أصغر من القيمة الحرجة, لذلك نقبل فرضية العدم ,التي تقول بعدم بوجود توافق معنوي بين هذين المتحولين.

وبناءً على ذلك يمكننا حساب قيمة معامل التوافق C (حسب علاقة بيرسون) فنجد أن:

$$C_1 = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{3,59}{3,59 + 192}} = \sqrt{0,0184} = 0,135$$

وهي قيمة صغيرة تدل على عدم وجود توافق بين الحالة الاجتماعية والجنس في العينة المسحوبة.

ملاحظة: كان يمكننا حساب معامل (كرامر) حيث نجد أن:

$$C_1 = \sqrt{\frac{\chi^2}{n(q-1)}} = \sqrt{\frac{3,59}{192(2-1)}} = 0,1367$$

وهي قيمة قريبة من المعامل السابق، وتدل أيضاً على عدم وجود توافق بين الحالة الاجتماعية والجنس.

10-3: اختبار استقلالية متحولين مرتبين بواسطة χ^2 :(من عينة واحدة)

إن دراسة استقلالية متحولين نوعيين (أو كميين) بوساطة χ^2 تقتضي أن يأخذ كل منهما أو أحدهما عدة حالات مرتبة أو مبوبة في جدول مزدوج يسمى جدول التوافق من المرتبة $(k \times \ell)$، وهو يتضمن التكرارات المطلقة في كل خلية مقابلة لتقاطع كل حالتين لهما. ولتوضيح ذلك نفترض أنه لدينا متحولين Y, X، ومأخوذين من عينة واحدة, وأن المتحول X يأخذ K حالة متنافية ومرتبة هي:

$$X : (x_1 \quad x_2 \quad x_3 \quad x_i .... x_k)$$

وأن المتحول Y يأخذ ℓ حالة متنافية ومرتبة هي:

$$Y : (y_1 \quad y_2 \quad y_3 \quad y_j .... x_\ell)$$

وبما أن المؤشر χ^2 يعرف على التكرارات الفعلية والتكرارات المتوقعة (النظرية)، فإننا نرمز للتكرارات الفعلية المقابلة للحالة المزدوجة (x_i, y_i) بالرمز n_{ij}، وللتكرارات المتوقعة المقابلة لنفس الحالة (x_i, y_i) بالرمز m_{ij}. ولنفترض أن التكرارات المقابلة لتقاطعات هذه الحالات المرتبة محسوبة وموجودة في الجدول المزدوج التالي:

جدول(10-7): التكرارات الفعلية والمتوقعة المقابلة لتقاطعات حالات المتحولين

حالات X حالات Y	x_1	x_2	x_3	....	x_i	....	x_k	$\sum$
Y_1	n_{11} m_{11}	n_{21} m_{21}	n_{31} m_{31}	....	n_{i1} m_{i1}	....	n_{k1} m_{k1}	n'_1 m'_1
Y_1	n_{12} m_{12}	n_{22} m_{22}	n_{32} m_{32}	....	n_{i2} m_{i2}	....	n_{k2} m_{k2}	n'_2 m'_2
......	....	....	....	....	....	....	....	....

Y_j	n_{1j} m_{1j}	n_{2j} m_{2j}	n_{3j} m_{3j}	....	n_{ij} m_{ij}	....	n_{kj} m_{kj}	n'_j m'_j
....	....	....	....	....	....	....	....	....
Y_ℓ	$n_{1\ell}$ $m_{1\ell}$	$n_{2\ell}$ $m_{3\ell}$	$n_{3\ell}$ $m_{3\ell}$	....	$n_{i\ell}$ $m_{i\ell}$	....	$n_{k\ell}$ $m_{k\ell}$	n'_ℓ m'_ℓ
$\sum$	n_1	n_2	n_3	....	n_i	....	n_k	N

وهنا نلاحظ أنه قد رمزنا لمجموع التكرارات في العمود i بالرمز n_i، ولمجموع التكرارات في السطر j بالرمز n'_j

أي أن: $n'_j = \sum_{i=1}^{k} n_{ij}$ ، $n_i = \sum_{j=1}^{\ell} n_{ij}$

ورمزنا لمجموع التكرارات في جميع الخلايا (الأسطر والأعمدة) بـ n.

أي أن: $n = \sum_{j=1}^{\ell} \sum_{i=1}^{k} n_{ij} = \sum_{i=1}^{k} n_i = \sum_{j=1}^{\ell} n'_j$

ولحساب التكرارات المتوقعة m_{ij} نستخدم طريقة التناسب، التي تقتضي أن يكون عدد التكرارات المتوقعة (مثلاً) في الخلية (1، 1) جزءاً من المجموع n'_1 ومتناسباً مع حصتها من مجموع التكرارات n؛ أي أنه يجب أن يكون متناسباً مع الكسر $\left(\frac{n_1}{n}\right)$، وبذلك نجد أن التكرار المتوقع m_{11} يحسب من العلاقة:

$$m_{11} = \frac{n_1}{n} \cdot n'_1$$

ويمكن حسابها كجزء من n_1، متناسب مع حصتها $\frac{n'_1}{n}$, فيكون $m_{11} = \frac{n'_1}{n} \cdot n_1$

وفي كلتا الحالتين يكون لدينا:

$$m_{11} = \frac{n_1 \cdot n'_1}{n} = \frac{\text{مجموع السطر الأول × مجموع العمود الأول}}{\text{المجموع الكلي } n} \qquad (8-10)$$

وبصورة عامة تحسب التكرارات المتوقعة m_{ij} في جميع خلايا الجدول المزدوج من العلاقة:

$$m_{ij} = \frac{n_i \cdot n'_j}{n} = \frac{\text{المجموع في السطر j × المجموع في العمود i}}{\text{المجموع الكلي للتكرارات}} \qquad (9-10)$$

وأخيراً نجد أن المؤشر χ^2 يعرف أيضاً بالعلاقة الآتية:

$$\chi^2 = \sum_{j=1}^{\ell} \sum_{i=1}^{k} \left[\frac{\left(n_{ij} - m_{ij}\right)^2}{m_{ij}} \right] \qquad (10-10)$$

ولقد أشرنا سابقاً إلى أن المتحول χ^2 يخضع للتوزيع الاحتمالي $\chi^2(x)$ بدرجة حرية تساوي $v = (k-1)(\ell-1)$.

شروط تطبيق المؤشر χ^2 :

لا يجوز تطبيق المؤشر χ^2 إلا ضمن الشروط التالية:

1. فرضية العدم أو الاستقلال تكون كما يلي: إن المتحولين المدروسين $(x \;,\; y)$ مستقلين عن بعضهما البعض، أي عدم وجود علاقة بين المتحولين المدروسين.
2. الفرضية البديلة تكون كما يلي: إن المتحولين المدروسين غير مستقلين وبالتالي فهما مرتبطان إحصائياً.
3. أن تكون العينة المسحوبة للدراسة هي عينة عشوائية وأن لا يقل حجمها عن 50 وحدة, وتشكل اشباعاً لخلايا الجدول المتقاطعة.
4. أن لا تكون قيمة أي من قيم التكرارات المتوقعة في أي خلية أقل من 5، لأن ذلك يؤثر على عملية التقدير، ويمكن معالجة هذه الحالة عند ظهورها بأن نقوم بدمج عمودين متجاورين، أو سطرين متجاورين حتى نتخلص من تلك التكرارات القليلة. ويمكن أن نتساهل بهذا الشرط واستبداله بشرط آخر، وهو أن لا يزيد عدد الخلايا التي يكون تكرارها المتوقع أقل من 3 عن $\frac{1}{5}$ عدد الخلايا.
5. أن يتم اتخاذ القرار حول استقلال أو عدم استقلال المتحولين المدروسين بواسطة مقارنة القيمة المحسوبة للمؤشر χ^2 مع القيمة الجدولية لمتحول التوزيع χ^2_α المقابلة لمستوى الدلالة α ولدرجة حرية $v=(k-1)(\ell-1)$ ، ويتم اتخاذ القرار كما يلي:

. إذا كانت القيمة المحسوبة χ^2 أصغر أوتساوي من القيمة الجدولية χ^2_α نقبل فرضية العدم ونعتبر المتحولين مستقلين عن بعضهما البعض أي أنهما غير مرتبطين.

– . إذا كانت القيمة المحسوبة χ^2 أكبر من القيمة الجدولية χ^2_α نرفض فرضية العدم التي تقول باستقلال المتحولين، ونقبل الفرضية البديلة التي تقول بوجود علاقة ارتباطية بينهما، وكلما كان الفرق كبيراً كانت متانة الارتباط كبيرة.

مثال (10–5): لنفترض أننا قمنا بتبويب بيانات عينة عشوائية مؤلفة من مبيعات 160 سيارة حسب أنواعها ودرجة تعليم المشترين لها فحصلنا على الجدول التكراري التالي:

جدول(10–8): بيانات المثال لـ $n_{i\,j}$ المقابلة لخلايا المزدوجة:

أنواع السيارات درجة التعليم	فخمة	متوسطة	عادية	خفيفة	مجموع التكرارات الهامشية	التكرارات النسبية الهامشية
شهادة عليا	10	8	6	4	28	$\frac{28}{160}$
جامعية	15	20	12	5	52	$\frac{52}{160}$
ثانوية	20	15	10	2	47	$\frac{47}{160}$
إعدادية وأقل	5	10	15	3	33	$\frac{33}{160}$

مجموع التكرارات الهامشية (للأعمدة)	50	53	43	14	160	1

والمطلوب : دراسة ارتباط (أو استقلال) درجة تعليم المشتري بنوع السيارة التي يشتريها.

الحل: إن حل هذه المسألة يعتمد على فرضية الاستقلال والتي تعاكس الارتباط والتي تكتب كما يلي:

فرضية العدم: عدم وجود ارتباط بين درجة تعليم المشتري ونوع السيارة التي يشتريها؛ أي أنهما مستقلان .

الفرضية البديلة: يوجد ارتباط بين درجة تعليم المشتري ونوع السيارة التي يشتريها.

فإذا كانت درجة تعليم المشتري مستقلة عن نوع السيارة التي اشتراها، فإن ذلك يعني أن مجموع التكرارات الداخلية في كل عمود يجب أن يتوزع بما يتناسب مع التكرارات النسبية الهامشية المقابلة لدرجات التعليم.

أي أنه إذا كانت مبيعات السيارات الفخمة مستقلة عن درجة التعليم المشتري فإن إجمالي حجم مبيعاتها وهو (50) سيارة، يجب أن يتوزع على درجات التعليم بشكل يتناسب مع نسبة كل درجة في الدراسة، وهذا يعني أنه إذا كان حجم المبيعات مستقلاً عن الدرجات فإننا نتوقع أن تكون التكرارات النظرية كما يلي:

. أن يكون توقع عدد السيارات الفخمة التي يشتريها ذوي الشهادات العليا كما يلي :

$50 \times \frac{28}{160} = 8{,}75$ سيارة فخمة

. أن يكون توقع عدد السيارات الفخمة المتوقعة للجامعيين: $50 \times \frac{52}{160} = 16{,}25$ سيارة فخمة

. أن يكون توقع عدد السيارات الفخمة المتوقعة للثانويين: $50 \times \frac{47}{160} = 14{,}69$ سيارة فخمة

. أن يكون توقع عدد السيارات الفخمة المتوقعة للاعداديين: $50 \times \frac{33}{160} = 10{,}31$ سيارة فخمة

وبإتباع نفس الطريقة نقوم بحساب التكرارات المتوقعة لأنواع السيارات الأخرى، وذلك بتطبيق العلاقة العامة التالية:

$$m_{ij} = \frac{n_i \cdot n'_j}{n} \qquad (11-10)$$

حيث أن n_i هي المجاميع العمودية للتكرارات، n'_j هي المجاميع السطرية للتكرارات، و n المجموع الكلي لها, فنحصل على الجدول التالي:

جدول(10-9): التكرارات المتوقعة لخلايا الجدول m_{ij}

أنواع السيارات درجة التعليم	فخمة	متوسطة	عادية	خفيفة	المجموع n'_j
شهادة عليا	8,75	9,28	7,52	2,45	28
جامعية	16,25	17,23	13,98	4,55	52
ثانوية	14,69	15,57	12,63	4,11	47
إعدادية وأقل	10,31	10,93	8,87	2,89	33
مجموع التكرارات الهامشية	50	53	43	14	160

n_i					

وهنا يجب أن نشير إلى أن المجاميع الهامشية للأسطر والأعمدة للتكرارات المتوقعة يجب أن تساوي مقابلاتها في التكرارات الفعلية السابقة.

وقبل أن نتابع الحل نلاحظ أن جميع التكرارات المتوقعة في عمود السيارات الخفيفة أقل من 5، وهذا خلل في أحد شروط تطبيق المعيار χ^2 (الشرط 4)، لذلك نقوم بدمج عمود السيارات الخفيفة مع العمود المجاور له وهو عمود السيارات العادية، ونجمع التكرارات الفعلية لهما، كما نجمع التكرارات المتوقعة لهما، ونضع النتائج في جدول موحد، فنحصل على الجدول التالي:

جدول(10–10): التكرارات الفعلية والمتوقعة بعد الدمج

أنواع السيارات / مهنة المشتري	فخمة	متوسطة	عادية وخفيفة	المجموع n'_j
شهادة عليا	10	8	10	28
	8,75	9,28	9,79	
جامعية	15	20	17	52
	16,25	17,23	18,53	
ثانوية	20	15	12	47
	14,69	15,57	16,74	
إعدادية وأقل	5	10	18	33
	10,31	10,93	11,76	
المجموع n_i	50	53	57	160

وهكذا يصبح الجدول محققاً لجميع شروط تطبيق المعيار χ^2 , لذلك نقوم بحساب الحدود المختلفة للتركيب $\frac{(n_{ij}-m_{ij})^2}{m_{ij}}$ ونضعها في الجدول التالي:

جدول(10–11): قيم الحدود $\frac{(n_{ij}-m_{ij})^2}{m_{ij}}$

أنواع السيارات / مهنة المشتري	فخمة	متوسطة	عادية وخفيفة	المجموع
شهادة عليا	0,17857	0,17655	0,00009	
جامعية	0,09615	0,44532	0,12633	
ثانوية	1,91941	0,020867	1,34215	
إعدادية وأقل	2,73483	0,07913	3,31102	
مجموع التكرارات الهامشية	4,92896	0,72187	4,77959	10,43043= x^2

ومن الجدول السابق نجد أن قيمة χ^2 المحسوبة $\chi^2 = 10,43042$

ومن الجداول الملحقة نجد أن قيمة χ^2 الحرجة المقابلة لمستوى دلالة $\alpha = 0,05$ ودرجة حرية $v = (k-1)(\ell-1) = (4-1)(3-1) = 6$, تساوي $\chi^2_{\alpha 6} = 12,59$. وبمقارنة هاتين القيمتين نجد أن

قيمة χ^2 المحسوبة أصغر من $\chi^2_{\alpha 6}$ ، لذلك نقبل فرضية العدم، والتي تقول بأن المتحولين مستقلان، ونستنتج أن درجة تعليم المشتري لا ترتبط بنوع السيارة التي يشتريها(بعد دمج الخفيفة مع العادية).

10-4: الاختبارات اللامعلمية لتطابق التوزيعات الاحتمالية للمتحولات الكمية أو المرتبة (من عينة واحدة):

تتميز المتحولات الكمية والرتبية بإمكانية ترتيب حالاتها المختلفة حسب علاقة ترتيب معينة مثل: أكبر، أو أصغر، أو أفضل...إلخ. وبذلك تأخذ هذه الحالات اتجاهاً معيناً (متصاعداً، أو متنازلاً) وتكون مرفقة بالتكرارات المطلقة (أو النسبية) المقابلة لكل حالة فيها. وعندها نريد اختبار توافق تلك التكرارت النسبية مع أحد التوزيعات الاحتمالية المعروفة، وأهم الاختبارات المعرفة على تلك التكرارات الاختبارات التالية:

10-4-1: اختبارات تطابق التكرارات النسبية لقيم متحول X مع توزيع احتمالي معين:

لنفترض أن متحولاً كمياً X له الحالات والتكرارات المطلقة والنسبية التالية:

الحالات المرتبة للمتحول x	x_1	x_2	x_3	…	x_i	…	x_m	المجموع
n_i التكرارات المطلقة	n_1	n_2	n_3	…	n_i	…	n_m	n
p_i التكرارات النسبية	p_1	p_2	p_3	…	p_i	…	p_m	1

حيث إن p_i تحسب من العلاقات النسبية التالية: $p_i = \frac{n_i}{n}$

ولنفترض أننا نريد دراسة مدى تطابق أو توافق هذه التكرارات النسبية مع قيم أحد التوزيعات الاحتمالية النظرية المعطاة بدلالة علاقة رياضية معينة مثل $f\left(x_i\right)$ ، وبذلك يأخذ التوزيع النظري، أو المتوقع الشكل الآتي:

الحالات المرتبة للمتحول X	x_1	x_2	x_3	…	x_i	…	x_m	المجموع
الاحتمالات النظرية المقابلة	$f\left(x_1\right)$	$f\left(x_2\right)$	$f\left(x_3\right)$	…	$f\left(x_i\right)$	…	$f\left(x_m\right)$	1

ولدراسة مدى التطابق بين هذه التكرارات النسبية مع التوزيعات الاحتمالية نستخدم أحد الاختبارات الآتية:

أ . اختبار χ^2:

وهو الاختبار الذي درسناه في الفقرة السابقة، والذي يعرف في حالة متحول واحد بالعلاقة التالية:

$$\chi^2 = \sum_{i=1}^{m} \frac{\left(n_i - m_i\right)^2}{m_i} = n \sum_{i=1}^{m} \frac{\left(p_i - f x_i\right)^2}{f\left(x_i\right)} \qquad (12-10)$$

مثال (10-6): حول المطابقة مع التوزيع الطبيعي:

لنفترض أن دراسة لأعمار المسافرين عبر أحد المعابر الحدودية شملت 100 مسافر, قامت بتبويب أفراد هذه العينة حسب العمر إلى 6 فئات عشرية كان كما يلي:

جدول(10−12): بيانات المثال

رقم الفئة i	1	2	3	4	5	0	
الفئة العمرية	[0-10[	[10-20[	[20-30[	[30-40[	[40-50[	[50-60]	$\sum$
التكرارات المطلقة	5	20	15	45	10	5	100
مركز الفئة	5	15	25	35	45	55	-

والمطلوب بيان فيما إذا كان توزيع أعمار المسافرين يتطابق مع التوزيع الطبيعي، ثم اختبار ذلك عند مستوى دلالة $\alpha = 0{,}05$.

الحل: نضع الفرضتين كما يلي:

H_0: إن توزيع أعمار المسافرين يخضع للتوزيع الطبيعي الذي لا نعلم توقعه ولا انحرافه المعياري.

H_1: إن توزيع أعمار المسافرين لا يخضع للتوزيع الطبيعي.

ولتحديد التوزيع الطبيعي المفترض نقوم بحساب متوسط العينة $\bar{x}$ ونعتبره تقديراً لمتوسط المجتمع μ ، ثم نقوم بحساب تباين العينة s^2 ونعتبره تقديراً لتباين المجتمع σ^2، ثم نعوضهما في معادلة التوزيع الطبيعي العام فنحصل على الصيغة المحددة للتوزيع الطبيعي المفروض. وبناء على ذلك نجد أن متوسط العينة يساوي:

$$\bar{x} = \frac{1}{n}\sum n_i x_i = \frac{1}{100}\left[5\cdot 5 + 20\cdot 15 + 15\cdot 25 + 45\cdot 35 + 10\cdot 45 + 5\cdot 55\right]$$

$$\bar{x} = \frac{1}{100}(3000) = 30 \quad \text{عاماً}$$

وأن تباين العينة يساوي:

$$s^2 = \frac{1}{n-1}\sum n_i\left(x_i - \bar{x}\right)^2 = \frac{1}{99}\sum n_i\left(x_i - 30)^2\right)$$

$$s^2 = \frac{1}{99}\begin{bmatrix} 5(5-30)^2 + 20(15-30)^2 + 15(25-30)^2 + \\ + 45(25-30)^2 + 10(25-30)^2 + 5(25-30)^2 \end{bmatrix}$$

$$s^2 = \frac{1}{99}(3125 + 4500 + 375 + 1125 + 1125 + 2250 + 3125)$$

$$s^2 = \frac{1}{99}(14500) = 1465{,}46$$

وبذلك نجد أن الانحراف المعياري للعينة s يساوي:

$$s = \sqrt{146{,}46} = 12{,}1$$

وهكذا يمكننا تقدير متوسط المجتمع وانحرافه المعياري من العلاقتين:

$$\tilde{\mu} = \bar{x} = 30$$

$$\tilde{\sigma} = s = 12{,}1$$

وبعدها نعوض ذلك في معادلة التوزيع المفترض, فنجد أن معادلة التوزيع الطبيعي المفترض تأخذ الشكل الآتي:

$$f\ (x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{-1}{2}\left(\frac{x-\bar{x}}{\sigma}\right)^2}$$

$$f\ (x) = \frac{1}{12{,}1\sqrt{2\pi}} e^{-\frac{-1}{2}\left(\frac{x-30}{12{,}1}\right)^2} \qquad (13-10)$$

واعتماداً على هذا التوزيع المفترض نقوم بحساب الاحتمالات النظرية المقابلة للفئات العمرية السابقة، فنجد أن:

$$\begin{aligned} P(0 \le x < 10) &= \Phi\left(\frac{10-30}{12{,}1}\right) - \Phi\left(\frac{0-30}{12{,}1}\right) \\ &= \Phi(-1{,}65) - \Phi(-2{,}48) \\ &= 1 - \Phi(1{,}65) - [1 - \Phi(2{,}48)] \\ &= \Phi(2{,}48) - \Phi(1{,}65) \\ &= 0{,}9934 - 0{,}9505 \\ &= 0{,}0429 \end{aligned}$$

وكذلك نجد أن:

$$\begin{aligned} P(10 \le x < 20) &= \Phi\left(\frac{20-30}{12{,}1}\right) - \Phi\left(\frac{10-30}{12{,}1}\right) \\ &= 1 - \Phi(0{,}83) - [1 - \Phi(-1{,}65)] \\ &= 1 - 0{,}7967 - (1 - 0{,}9505) = 0{,}1538 \end{aligned}$$

$$\begin{aligned} P(20 \le x < 30) &= \Phi\left(\frac{30-30}{12{,}1}\right) - \Phi\left(\frac{20-30}{12{,}1}\right) \\ &= 0{,}5 - (1 - 0{,}7967) = 0{,}2967 \end{aligned}$$

$$\begin{aligned} P(30 \le x < 40) &= \Phi\left(\frac{40-30}{12{,}1}\right) - \Phi\left(\frac{30-30}{12{,}1}\right) \\ &= 0{,}7967 - 0{,}5 = 0{,}2967 \end{aligned}$$

$$\begin{aligned} (40 \le x < 50) &= \Phi\left(\frac{50-30}{12{,}1}\right) - \Phi\left(\frac{40-30}{12{,}1}\right) \\ &= 0{,}9505 - 0{,}7967 = 0{,}1538 \end{aligned}$$

$$\begin{aligned} P(50 \le x < 60) &= \Phi\left(\frac{60-30}{12{,}1}\right) - \Phi\left(\frac{50-30}{12{,}1}\right) \\ &= 0{,}9934 - 0{,}9605 = 0{,}0429 \end{aligned}$$

ولحساب قيمة المؤشر χ^2 نضع تلك الاحتمالات في جدول كالآتي ونضربها بحجم العينة $n = 100$ فنحصل على التكرارات المتوقعة حسب التوزيع الطبيعي المفروض كما يلي.

جدول(10-13): التكرارات المتقابلة مع الحسابات اللازمة

I	1	2	3	4	5	6	$\sum$
Xi	0-10	10-20	20-30	30-40	40-50	50-60	-
ni	5	20	15	45	10	5	100
P_i	0,0429	0,1538	0,2967	0,2967	0,1538	0,0429	1
$m_i = nP_i$	4,29	15,38	29,67	29,67	15,38	4,29	100
$(n_i - m_i)^2$	0,5041	21,34	215,21	235,0	28,94	0,5041	–
$\frac{(n_i - m_i)^2}{mi}$	0,1175	1,3878	7,253	7,9207	1,8820	0,1175	$18,6785 = \chi^2$

وبذلك نجد أن قيمة χ^2 الفعلية أو المحسوبة تساوي $\chi^2 = 18,6785$.

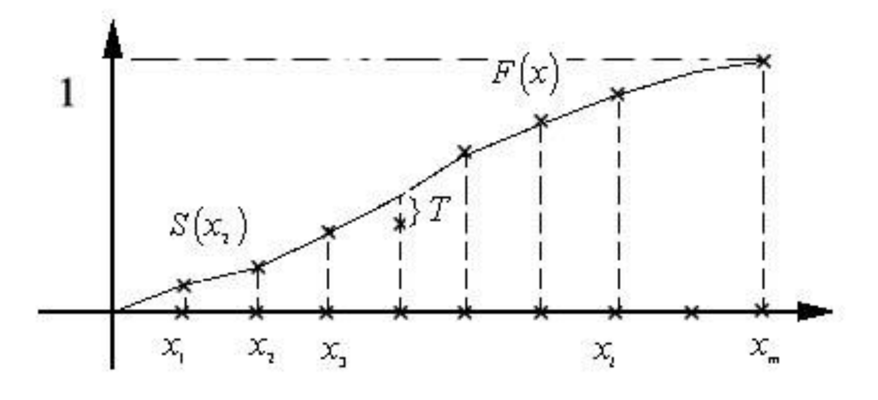
ومن الجداول الملحقة نجد أن قيمة χ^2 الجدولية المقابلة لدرجة حرية مساوية لـ $k - 2 - 1 = 3$ (نقصت درجة الحرية بمقدار اثنين لأننا استفدنا من بيانات العينة في تقدير متوسط المجتمع μ وانحرافه المعياري σ) ولمستوى دلالة $\alpha = 0,05$ تساوي $\chi^2_{0,05} = 7,815$, وبالمقارنة نجد أن قيمة χ^2 المسحوبة أكبر من قيمة χ^2_α الجدولية، لذلك نرفض فرضية العدم H_0, ونقبل الفرضية البديلة التي تقول أن توزيع عمر المسافر لا يخضع للتوزيع الطبيعي الذي قدرنا متوسطه بـ 30 عاماً، وانحرافه المعياري بـ 12,1. ومع ذلك فهو قد يكون خاضعاً لتوزيع طبيعي آخر بمتوسط آخر وانحراف معياري آخر.

ب . اختبار (كولموغوروف . سميرنوف Kolmogorov-Smirnov):

وهو يعتمد على مقارنة تابع التوزيع الاحتمالي التجريبي $S(x_i)$ (التراكمي) مع تابع التوزيع النظري $F(x_i)$ (التراكمي). وإذا حسبنا هذين التابعين ووضعناهما في جدول نحصل على ما يلي:

الحالات المرتبة للمتحول x	X_1	X_2	X_3	...	X_i	...	X_m	المجموع
تابع التوزيع التجريبي $S(x_i)$	$S(x_1)$	$S(x_2)$	$S(x_3)$	...	$S(x_i)$	...	1	
تابع التوزيع النظري $F(x_i)$	$F(x_1)$	$F(x_2)$	$F(x_3)$		$F(x_i)$		1	

ويمكن للشكل البياني لهذين التابعين أن يأخذ الشكل التالي:



وعندها يمكننا أن نضع الفرضيتين كما يلي:

$H_0 : S(x_i) = F(x_i) \quad \forall i$

$H_1 : S(x_i) \neq F(x_i)$ (حالة ثنائي الجانب فقط)

وبناءً على ذلك قام (كولموغوروف وسميرنوف) بتعريف مؤشر اختبارهما بالعلاقة التالية:

$T = \sup|S(x_i) - F(x_i)| =$ أكبر قيمة للفرق بالقيمة المطلقة

ثم قاما بإعداد جداول خاصة للتوزيع الاحتمالي الذي يخضع له المؤشر T، وهي ملحقة في آخر الكتاب، ومنها يمكننا الحصول على القيم الحرجة T^* المقابلة لمستوى الدلالة α ولحجم العينة n. ولاتخاذ قرار حول الفرضية H_0 نقارن قيمة T المحسوبة مع القيمة الحرجة T^*. فإذا كانت $T < T^*$ نقبل فرضية العدم H_0، التي تقول بعدم وجود فرق بين التوزيعين, والعكس بالعكس.

مثال (10-7): قمنا بتوبيب أفراد عينة من العمال حجمها $n = 25$ فرداً حسب الحالة التعليمية، فحصلنا على الجدول التالي:

جدول(10-14): بيانات المثال

الحالات التعليمية لـ X	أمي	ابتدائي	إعدادي	ثانوي	المجموع
التكرارات المطلقة	10	8	5	2	25
التكرارات النسبية	$\frac{10}{25}$	$\frac{8}{25}$	$\frac{5}{25}$	$\frac{2}{25}$	1

فهل هذا التوزيع يخضع للتوزيع المنتظم. اختبر بواسطة اختبار (كولموغوروف وسميرنوف) بمستوى دلالة $\alpha = 0{,}10$.

الحل: بما أن المطلوب مقارنة هذا التوزيع بالتوزيع المنتظم فإن قيم التوزيع الاحتمالي النظري تأخذ الشكل التالي:

جدول(10-15) : قيم التوزيع المنتظم

الحالات التعليمية لـ X	أمي	ابتدائي	إعدادي	ثانوي	المجموع
قيم التوزيع النظري	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	1

ولمقارنة هذين التوزيعين نقوم بوضع الفرضيتين كما يلي:

$H_0 : S(x_i) = F(x_i) \quad \forall i$

$H_1 : S(x_i) \neq F(x_i)$

ثم نقوم بحساب قيم تابعي التوزيعين التجريبي والنظري (تراكمياً) فنحصل على الجدول التالي:

جدول (16a–10): القيم التراكمية لتابعي التوزيع

الحالات التعليمية لـ X	أمي	ابتدائي	إعدادي	ثانوي
قيم تابع التوزيع التجريبي $S(x)$	$\frac{10}{25}$	$\frac{18}{25}$	$\frac{23}{25}$	1
قيم تابع التوزيع النظري $F(x)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{3}{4}$	1
$S(x_i)-F(x_i)$	0,15	0,22	0,17	0

ثم نقوم بحساب الفروقات $S(x_i)-F(x_i)$ ونضيفها إلى الجدول السابق، فنحصل على السطر الأخير فيه، وبذلك نجد أن قيمة مؤشر الاختبار T تساوي:

$$T=\sup|S(x_i)-F(x_i)|=0,22$$

ومن الجدول الخاص الملحق في آخر الكتاب نجد أن القيمة الحرجة لـT المقابلة لمستوى دلالة $\alpha=0,10$ ولـ $n=25$ تساوي $T^*=0,238$. وبالمقارنة نجد أن $T<0,238$ لذلك نقبل فرضية العدم H_0، ونعتبر أن التوزيع التجريبي يخضع أو يتقارب مع التوزيع المنتظم.

جـ . اختبار ليليفورز Lilliefors) للتطابق مع التوزيع الطبيعي المعياري:

يطبق هذا الاختبار لدراسة التطابق بين التوزيع التجريبي لمتحول كمي X مع التوزيع الطبيعي المعياري $N(0\ ,\ 1)$. ولنفترض أنه لدينا عينة بحجم n ومبوبة في جدول كالتالي:

جدول(10–16):البيانات الإحصائية

قيم X المرتبة	X_1	X_2	X_3	...	X_i	...	X_m	المجموع
التكرارات المطلقة	n_1	n_2	n_3	...	n_i	...	n_m	n
التكرارات النسبية	p_1	p_2	p_3	...	p_i	...	p_m	1
تابع التوزيع التراكمي $S(x_i)$	$S(x_1)$	$S(x_2)$	$S(x_3)$	...	$S(x_i)$	...	1	

ومن هذه البيانات نقوم بتقدير متوسط المجتمع μ وتباينه σ^2 من العلاقتين:

$$\tilde{\mu}=\bar{x}=\frac{1}{n}\sum n_i x_i$$

$$\tilde{\sigma}^2=s^2=\frac{1}{n-1}\sum n_i(x_i-\bar{x})^2$$

وبعدها نقوم بحساب القيم المعيارية للمتحول Z من العلاقة:

$$Z_i=\frac{x_i-\bar{x}}{s} \qquad i:1\ ,\ 2\ ,\ 3......m$$

ونضع الفرضتين كما يلي:

فرضية العدم H_0. المتحول Z يخضع للتوزيع الطبيعي المعياري $N(0\ ,1)$، وبالتالي يكون المتحول X خاضعاً للتوزيع الطبيعي العام $N(\mu\ ,\ \sigma^2)$.

الفرضية البديلة: المتحول Z لا يخضع للتوزيع الطبيعي المعياري $N(0\,,1)$.

ثم نقوم بتعريف مؤشر الاختبار على المتحول Z الذي له تابع التوزيع التجريبي نفسه $S(z_i)$، ومقارنته مع تابع التوزيع الطبيعي المعياري التراكمي $\Phi(x_i)$ (المعرف سابقاً)، وذلك باستخدام مؤشر (كولموغوروف - سميرنوف) نفسه الذي يأخذ الشكل التالي:

$$T_z = \sup\left|S(z_i) - \Phi(z_i)\right|$$

ولاتخاذ قرار حول الفرضية H_0 نقارن قيمة T_z مع القيمة الحرجة لها T_z^* المأخوذة من الجداول الخاصة بذلك والمقابلة لمستوى الدلالة α ولحجم العينة n. فإذا كانت $T_z < T_z^*$، نقبل الفرضية H_0، والعكس بالعكس.

مثال (10-8):

لنفترض أن تبويب الطلاب المبحوثين حسب العمر في عينة بحجم $n = 50$ ، كان كما يلي:

جدول(10-17): بيانات المثال

قيم العمر X	20	21	22	23	24	25	26	المجموع
التكرارات المطلقة	5	7	8	10	9	8	3	50
التكرارات النسبية	$\frac{5}{50}$	$\frac{7}{50}$	$\frac{8}{50}$	$\frac{10}{50}$	$\frac{9}{50}$	$\frac{8}{50}$	$\frac{3}{50}$	1
التكرارات النسبية التراكمية $S(x_i)$	$\frac{5}{50}$	$\frac{12}{50}$	$\frac{20}{50}$	$\frac{30}{50}$	$\frac{39}{50}$	$\frac{47}{50}$	1	

والمطلوب دراسة فيما إذا كان العمر X يخضع للتوزيع الطبيعي العام، واختبار ذلك بمستوى دلالة $\alpha = 0{,}05$.

الحل: نضع الفرضتين كما يلي:

H_0: أعمار المبحوثين تخضع للتوزيع الطبيعي العام $N(\mu, \sigma^2)$.

H_1: أعمار المبحوثين لا تخضع للتوزيع الطبيعي العام.

لذلك نقوم بتقدير كل من μ و σ^2 من العلاقتين:

$$\tilde{\mu} = \bar{x} = \frac{1}{50}\left[5\cdot 20 + 7\cdot 21 + 8\cdot 22 + 10\cdot 23 + 9\cdot 24 + 8\cdot 25 + 3\cdot 26\right]$$

$$\tilde{\mu} = \bar{x} = \frac{1147}{50} = 22{,}94$$

$$\tilde{\sigma}^2 = s^2 = \frac{1}{49}[5\cdot(20-22{,}94)^2 + 7\cdot(21-22{,}94)^2 +$$
$$8\cdot(22-22{,}94)^2 + 10\cdot(23-22{,}94)^2 + 9\cdot(24-22{,}94)^2 +$$
$$8\cdot(25-22{,}94)^2 + 3\cdot(26-22{,}94)^2]$$

$$s^2 = \frac{1}{49}[43{,}218 + 26{,}345 + 7{,}069 + 0{,}036 + 10{,}112 + 33{,}95 + 28{,}091]$$

$$s^2 = \frac{1}{49}(148{,}821) = 3{,}037 \qquad s = 1{,}743$$

ومنها نجد أن الانحراف المعياري يقدر بـ $\tilde{\sigma} = s = 1{,}743$.

ثم نقوم بحساب قيم المتحول المعياري z من العلاقة: $z_i = \frac{x_i - \bar{x}}{s}$ ، ونضعها مع التكرارات النسبية التراكمية في جدول كالتالي:

جدول(10−18): التكرارات النسبية التراكمية والاحتمالات المعيارية التراكمية

قيم z المعيارية	-2,94	-1,94	-2,94	0,06	1,06	2,06	3,06
التكرارات النسبية التراكمية $S(x_i)$	$\frac{5}{50}$	$\frac{12}{50}$	$\frac{20}{50}$	$\frac{30}{50}$	$\frac{39}{50}$	$\frac{47}{50}$	1
قيم التوزيع التجريبي التراكمية $S(x_i)$	$\frac{5}{50}$	$\frac{12}{50}$	$\frac{20}{50}$	$\frac{30}{50}$	$\frac{39}{50}$	$\frac{47}{50}$	
قيم التوزيع الطبيعي التراكمية $\Phi(x_i)$	0,0164	0,0762	0,1736	0,5239	0,8554	0,9803	1
الفروقات $\lvert\Phi(x_i) - F(x_i)\rvert$	0,0836	0,2138	0,2264	0,0761	0,0754	0,0403	0

ولقد قمنا بحساب الاحتمالات التراكمية النظرية $\Phi(x_i)$ اعتماداً على جدول التوزيع الطبيعي المعياري، فوجدنا مثلاً أن:

$$\Phi(-2{,}94) = 1 - \Phi(2{,}94)) = 1 - 0{,}99835 = 0{,}0164$$

ثم حسبنا الفروقات بين كل احتمالين تراكميين، ووضعنا هذه النتائج في الجدول السابق.

ومن ذلك الجدول نجد أن قيمة مؤشر (ليليفورز) تساوي:

$$T_1 = \sup\lvert S(z_i) - \Phi(z_i)\rvert = 0{,}2264$$

ولاتخاذ قرار حول ذلك التطابق نبحث عن القيمة الحرجة T_1^* من الجدول الخاص بها والمقابلة لمستوى دلالة $\alpha = 0{,}05$ ولـ $n = 50$، فنجد أنها تساوي $T_1^* = \frac{0{,}886}{\sqrt{50}} = 0{,}1253$. وبالمقارنة نجد أن:

$T > T_1^*$، لذلك نرفض فرضية العدم والتي تقول إن التوزيع الاحتمالي لأعمار المبحوثين يخضع للتوزيع الطبيعي العام، ونقبل الفرضية البديلة H_1 التي تقول إن أعمار المبحوثين لا تخضع للتوزيع الطبيعي العام.

10-5 : اختبارات الارتباط بين متحولين رتبيين (x, y) (من عينة واحدة):

إن اختبارات الارتباط لا تطبق على المتحولات الاسمية لأنها لا تقبل الترتيب، ولذلك فإن اختبارات الارتباط تشترط أن يكون المتحولان رتبيين على الأقل (يمكن أن يكون مجالي وتحويله إلى رتبي)، ولهما حالات متقابلة مكانياً أو زمانياً وعددها n زوجاً, وتؤخذ هذه الحالات من عناصر العينة المسحوبة. ولمعالجة هذه الاختبارات نفترض أنه لدينا متحولان رتبيان (x, y)، ولهما الحالات المتقابلة التالية:

$$X : x_1 \ , \ x_2 \ , \ x_3........x_n$$

$$Y : y_1 \ , \ y_2 \ , \ y_3........y_n$$

ونضع الفرضيتين كما يلي:

H_0: المتحولان X و Y مستقلان أو عدم وجود ارتباط بينهما.

H_1: المتحولان X و Y مرتبطان. (ويكون الاختبار ثنائي الجانب لأن الارتباط يمكن أن يكون موجباً، أو سالباً).

وإن أهم مقاييس الارتباط بين المتحولات الرتيبة هي:

أ- معامل الارتباط الرتبي (معامل سبيرمان):

ويعرف هذا المعامل بالعلاقة التالية:

$$RS = 1 - \frac{6\sum_{i=1}^{n}\left(r_i - q_i\right)^2}{n\left(n^2 - 1\right)} \qquad (14-10)$$

حيث إن r_i هي رتب المتحول الأول X. وإن q_i هي رتب المتحول الثاني Y.

و يبرهن في كتب مبادئ الإحصاء أن هذا المعامل هو حالة خاصة من معامل الارتباط العادي (معامل بيرسون)، لذلك فإنه يتمتع بجميع خواصه، وأهمها إنه محصور في المجال $[+1 \ , \ -1]$، وكلما كانت قيمته المطلقة $|RS|$ قريبة من الواحد، كان الارتباط قوياً.

ويمكن اختبار معنوية هذا المعامل باستخدام مؤشر التوزيع الطبيعي المعياري التالي:

$$z = \frac{RS - 0}{\frac{1}{\sqrt{n-1}}} = RS \cdot \sqrt{n-1} \qquad (15-10)$$

وبمقارنة قيمة Z المحسوبة مع قيمة $Z_{1-\frac{\alpha}{2}}$ الجدولية نتخذ القرار المناسب حول معنوية قيمة RS .

ولتطبيق هذا المعامل يجب القيام بالإجراءات التالية:

1 . ترتيب حالات المتحول الأول X تصاعدياً (أم تنازلياً) بوضع رتبة لكل حالة من حالاته وإعطائها رقماً صحيحاً من 1 حتى n. ونرمز لهذه الرتب بـ r_i.

2 . ترتيب حالات المتحول الثاني Y تصاعدياً (أم تنازلياً) بوضع رتبة لكل حالة من حالاته وإعطائها رقماً صحيحاً من 1 حتى n. ونرمز لهذه الرتب بـ q_i.

وهنا نشير إلى أن ترتيب حالات كل متحول يتم بشكل مستقل عن الآخر، ولكن بالاتجاه نفسه (تصاعدياً، أم تنازلياً).مع الحفاظ على الازواج المتقابلة لهما.

3 . نقوم بحساب الفروقات بين الرتب المتقابلة $(r_i - q_i)$، ثم نربعها $(r_i - q_i)^2$، ثم نأخذ مجموع تلك المربعات $\sum (r_i - q_i)^2$، ونعوضها في العلاقة (10−15) لحساب قيمة معامل الارتباط الرتبي RS.

مثال(10−9):

لدراسة العلاقة بين المستوى التعليمي للزوج، والمستوى التعليمي للزوجة، أخذنا عينة عشوائية بحجم $n = 7$ أسر، فحصلنا على البيانات التالية:

جدول(10−19): بيانات المثال

التسلسل	1	2	3	4	5	6	7
مستوى تعليم الزوج	أمي	ملم	ابتدائي	إعدادي	ثانوي	متوسط	جامعي
مستوى تعليم الزوجة	ملمة	أمية	إعدادية	ابتدائية	متوسط	ثانوية	ثانوية

والمطلوب دراسة الارتباط بين مستويي تعليم الزوج والزوجة.

الحل: بما أن مستويي التعليم متحولان رتبيان (قابلان للترتيب)، فإنه يمكننا تطبيق معامل الارتباط الرتبي عليهما لحساب متانة الارتباط بينهما، لذلك وقبل كل شيء نقوم بما يلي:

1 . إعطاء رتب تصاعدية متسلسلة لحالات تعليم الزوج من 1 حتى 7 كما هي موضحة في الجدول التالي:

2 . إعطاء رتب تصاعدية متسلسلة لحالات تعليم الزوجة من 1 حتى 7، علماً بأن كل من الحالتين الأخيرتين (ثانوية وثانوية) يمكن أن تأخذان 5 أو6، ولكن بما أنهما متكافئتان نعطي كلاهما الرقم المتوسط 5,5 كما هو موضح في الجدول.

3 . نحسب الفروقات $(r_i - q_i)$، ثم نربعها، ونأخذ مجموعها ونحسب قيمة معامل الارتباط الرتبي RS.

جدول(10−20): رتب مستويات التعليم للزوج وللزوجة

مستوى تعليم الزوج	أمي	ملم	ابتدائي	إعدادي	ثانوي	متوسط	جامعي
رتب تعليم الزوج r_i	1	2	3	4	5	6	7
مستوى تعليم الزوجة	ملمة	أمية	إعدادية	ابتدائية	متوسط	ثانوية	ثانوية
رتب تعليم الزوجة q_i	2	1	4	3	7	5,5	5,5
$(r_i - q_i)$	-1	$+1$	-1	$+1$	-2	$+0,5$	$+1,5$
$(r_i - q_i)^2$	1	1	1	1	4	0,25	2,25

وبذلك نجد أنّ $\sum(r_i - q_i)^2 = 10,5$

وأنّ معامل الارتباط الرتبي بين المستويين يساوي:

$$RS = 1 - \frac{6\sum_{i=1}^{n}(r_i - q_i)^2}{n(n^2-1)} = 1 - \frac{6\cdot(10,5)}{7\cdot(49-1)} \qquad (16\text{-}10)$$

$$RS = 1 - \frac{63}{336} = 0,8125 \qquad (17\text{-}10)$$

إن قيمة RS تدل على درجة ارتباط جيد بين مستويي التعليم للزوج والزوجة.

ولدراسة معنوية هذه القيمة نحسب قيمة مؤشر الاختبار Z فنجد أن:

$$Z = RS\cdot\sqrt{n-1} = 0,8125\sqrt{7-1} = 1,99 \qquad (18\text{-}10)$$

وبمقارنة القيمة المحسوبة 1.99 مع قيمة $Z_{1-\frac{\alpha}{2}}$ الجدولية المقابلة لمستوى دلالة $\alpha = 0,05$ والتي تساوي $Z_{0,975} = 1,96$ ، نجد أن $Z > 1,96$ ، لذلك نرفض فرضية العدم H_0 (التي تقول إن $RS = 0$) ونقبل الفرضية البديلة التي تقول إن قيمة RS هي قيمة معنوية وتدل على وجود ارتباط جيد بين المتحولين المدروسين.

ب : معامل ارتباط (كيندال Kendall**) من النوع** C:

ويستخدم هذا المعامل لدراسة متانة الارتباط بين متحولين رتيبين على الأقل، ولنفترض أن حالاتهما المأخوذة من عينة واحدة والمرتبة باتجاه واحد تتقاطع في جدول التكرارات كما يلي:

جدول(10−21): التكرارات المتقاطعة

حالات x المرتبة / حالات y المرتبة	x_1	x_2	x_3	x_4
y_1	n_{11}	n_{12}	n_{13}	n_{14}
y_2	n_{21}	n_{22}	n_{23}	n_{24}
y_3	n_{31}	n_{32}	n_{33}	n_{34}

ولا يشترط في هذه الحالة أن يكون عدد حالات X مساوياً لعدد حالات Y، ولكن يشترط أن تكون حالات كل منهما مرتبة ترتيباً باتجاه واحد(تصاعدياً، أو تنازلياً). ونرمز لعدد الأسطر بالرمز r، ولعدد الأعمدة بالرمز c، ولأصغرهما بالرمز q. ويعرف معامل كيندال من النوع c بالعلاقة التالية:

$$\tau_C = \frac{q(P-Q)}{N^2(q-1)}$$

حيث إن $q = \min(r,c)$

حيث أن: N هو حجم العينة المدروسة.

وأن P تحسب من اليسار إلى اليمين بأخذ جداء كل تكرار في مجموع التكرارات التي دونه (بعد حذف سطره وعموده) ثم أخذ مجموعها، أي إن:

$$P = n_{11}(n_{22}+n_{23}+n_{24}+n_{32}+n_{33}+n_{34}) + n_{12}(n_{23}+n_{24}+n_{33}+n_{34}) + n_{13}(n_{24}+n_{34}) + n_{21}(n_{32}+n_{33}+n_{34}) + n_{22}(n_{33}+n_{34}) + n_{23}(n_{34})$$

وأن Q تحسب من اليمين إلى اليسار بأخذ جداء كل تكرار في مجموع التكرارات التي دونه (بعد حذف سطره وعموده) ثم أخذ مجموعها؛ أي إن:

$$Q = n_{14}(n_{23}+n_{22}+n_{21}+n_{33}+n_{32}+n_{31}) + n_{13}(n_{22}+n_{21}+n_{32}+n_{31}) + n_{12}(n_{21}+n_{31}) + n_{24}(n_{33}+n_{32}+n_{31}) + n_{23}(n_{32}+n_{31}) + n_{22}(n_{31})$$

ملاحظة: إن المعامل τ_C يأخذ قيمه في المجال $[-1\,,+1]$، وكلما كانت قيمته المطلقة قريبة من الواحد كان الارتباط متيناً. ويمكن إجراء اختبار لمعنوية المعامل τ_C باستخدام مؤشر التوزيع المعياري التالي:

$$z_C = \frac{\tau_C - 0}{\sqrt{\dfrac{4(r+1)(c+1)}{9 \cdot n \cdot r \cdot c}}}$$

ثم اتخاذ القرار المناسب كالعادة.

مثال (10-10):

لدراسة العلاقة بين مستويي التعليم للزوج والزوجة، أخذنا عينة عشوائية كبيرة بحجم $n = 300$ أسرة، وبوبنا البيانات التالية:

جدول(10-22): بيانات المثال

حالات تعليم الزوج / حالات تعليم الزوجة	أمي	ابتدائي	إعدادي	ثانوي	جامعي	
أمية	20	10	15	12	-	
ابتدائية	10	15	20	10	5	
إعدادية	5	10	25	30	10	
ثانوية	-	3	15	40	45	
المجموع	35	38	75	92	60	300

والمطلوب دراسة الارتباط بين مستوى تعليم الزوج والزوجة.

الحل: نلاحظ أن المتحولين رتيبان، وأن حالاتهما مرتبة باتجاه تصاعدي. لذلك يمكننا حساب معامل كيندال من النوع c المعرف كما يلي:

$$\tau_C = \frac{q(P-Q)}{n^2(q-1)}$$

ولحساب P نأخذ جداء كل تكرار في مجموع التكرارات التي دونه (بعد حذف سطره وعموده) وذلك من اليسار إلى اليمين ثم نأخذ المجموع، فنجد أن:

$$\begin{aligned}P = {} & 20(15+20+10+5+10+25+30+10+3+15+40+45)+\\ & 10(20+10+5+25+30+10+15+40+45)+\\ & 15(10+5+30+10+40+45)+12(5+10+45)\\ & 10(10+25+30+10+3+15+40+45)+\\ & 15(25+30+10+15+40+45)+20(30+10+40+45)+10(10+45)\\ & 5(3+15+40+45)+10(15+40+45)+25(40+45)+30(45)\end{aligned}$$

وأخيراً نجد أن:

$$P = 21675$$

ولحساب Q نأخذ جداء كل تكرار في مجموع التكرارات التي دونه (بعد حذف سطره وعموده) وذلك من اليمين إلى اليسار، فنجد أن:

$$\begin{aligned}Q = {} & 12(20+15+10+25+10+5+15+3)+15(15+10+10+5)+\\ & 10(10+5)+5(30+25+10+5+40+15+3)+\\ & 10(25+10+5+15+3)+20(10+5+3)+15(5)+10(40+15+3)+\\ & 30(15+3)+25(3)=\end{aligned}$$

وأخيراً نجد أن:

$$Q = 4836$$

وبملاحظة أن $q=\min(4{,}5)=4$ وأن $n=300$، نجد أن قيمة معامل كيندال من النوع c تساوي:

$$\tau_C = \frac{4(21675-4836)}{(300)^2(4-1)} = 0{,}2495$$

وهذا يدل على وجود علاقة موجبة ولكنها ضعيفة بين مستويي تعليم الزوج والزوجة (حسب معطيات المسألة).

ولدراسة معنوية هذه القيمة نحسب مؤشر الاختبار Z_C، فنجد أن:

$$Z_C = \frac{\tau_C - 0}{\sqrt{\dfrac{4(r+1)(c+1)}{9\cdot n\cdot r\cdot c}}} = \frac{0{,}2495}{\sqrt{\dfrac{4\cdot 5\cdot 6}{9\cdot 300\cdot 4\cdot 5}}} = 5{,}2927$$

وبمـا أن قيمـة $Z_{1-\frac{\alpha}{2}}$ الجدوليـة المقابلـة $\alpha = 0,05$ تسـاوي 1,96، وأن $Z_C > 1,96$، لـذلك نـرفض فرضية العدم H_0 (التي تقول إن $\tau_C = 0$) ونعتبر أن القيمة التي حصلنا عليها معنوية، وتدل على وجود ارتباط معنوي بين مستويي تعليم الزوج والزوجة في العينة المدروسة.

ج : معامل الارتباط (غاما G) لمتحولين رتبيين (Goodman-Kruskal):

ويعرف هذا المعامل بالعلاقة التالية:

$$G = \frac{P - Q}{P + Q} \qquad (19-10)$$

حيث P و Q هما المعرفان بمعامل (كيندال) السابق، ويأخذ هذا المعامل قيمه في المجال $[-1 , +1]$، ويمكن اختبار معنوية القيمة التي نحصل عليها باستخدام مؤشر التوزيع الطبيعي المعياري المعرف بالعلاقة التالية:

$$Z_G = G \cdot \sqrt{\frac{P + Q}{\left[1 - G^2\right]}} \qquad (20-10)$$

مثال (10−11): لنأخذ المثال السابق ولنحسب المعامل RG لبياناته فنجد أن:

$$RG = \frac{21675 - 4836}{21675 + 4836} = \frac{16839}{25851} = 0,6352$$

وهو يدل على ارتباط جيد. ويمكن اختبار معنويته من خلال العلاقة (10− 20) السابقة.

10−6: اختبارات الارتباط بين متحول اسمي ومتحول رتبي(من عينة واحدة):

إن أهم هذه الاختبارات هي:

1 . معامل (كوريتون Coreton**):**

وهو يدرس العلاقة بين متحول رتبي مثل الحالة التعليمية، ومتحول اسمي ثنائي مثل الجنس (ذكر، أو أنثى)، أو مكان الإقامة (حضر، أو ريف)، أو الاستجابة (نعم، أم لا)..إلخ. ويعرف معامل (كوريتون) بالعلاقة التالية:

$$RC = \frac{2}{n}\left(\bar{y}_2 - \bar{y}_1\right)$$

حيث إن:

$\bar{y}_1$ هو متوسط رتب حالات المتحول الرتبي المقابلة للخاصة الأولى للمتحول الاسمي (للذكور، أو للحضر...).

$\bar{y}_2$ هو متوسط رتب حالات المتحول الرتبي المقابلة للخاصة الثانية للمتحول الاسمي (للإناث، أو للريف...).

n هو حجم العينة العشوائية.

ويأخذ هذا المعامل قيمه في المجال $[-1 \; , \; +1]$.

مثال (10−12):

ادرس العلاقة الارتباطية بين الحالة التعليمية والجنس (لعينة مؤلفة من عشرة أشخاص) وفق بيانات الجدول التالي:

جدول(10−23): بيانات المثال

رقم الشخص	1	2	3	4	5	6	7	8	9	10
الحالة التعليمية	أمي	إعدادي	ابتدائي	ثانوي	إعدادي	جامعي	إعدادي	أمي	ثانوي	ابتدائي
الجنس	ذكر	أنثى	أنثى	أنثى	ذكر	ذكر	أنثى	أنثى	ذكر	أنثى
رتب الحالة التعليمية لكل جنس	1	3	2	4	3	5	3	1	4	2

والمطلوب دراسة الارتباط بين الحالة التعليمية والجنس، حيث يتم ترتيب الحالات التعليمية لكلا الجنسين تصاعدياً من 1 للأمي وحتى 5 للجامعي.

الحل: نفترض أن حالة الذكر هي الصفة الأولى لمتحول الجنس، ونقوم بوضع رتب للذكور , وبذلك نجد أن متوسط الرتب التي تقابل حالة الذكر يساوي:

$$\bar{y}_1 = \frac{1+3+5+4}{4} = \frac{13}{4} = 3{,}25$$

ونفترض أن حالة الأنثى هي الصفة الثانية لمتحول الجنس، وبذلك نجد أن متوسط الرتب التي تقابل حالة الأنثى يساوي:

$$\bar{y}_2 = \frac{3+2+4+3+1+2}{6} = \frac{16}{6} = 2{,}5$$

وبذلك نجد أن معامل (كوريتون) بين التعليم والجنس يساوي:

$$RC = \frac{2}{10}\left(2,5-3,25\right) = -0{,}15 \qquad (21\text{-}10)$$

وهذا يدل على ارتباط عكسي ضعيف، ويميل لصالح الذكور ضد الإناث.

2: معامل ارتباط نقطة السلسلة المزدوجة (Point-Biserial coeff.):

ويطبق هذا المعامل على متحولين أحدهما كمي مجالي X ، والآخر اسمي ثنائي يأخذ القيمة 1 عندما تظهر الصفة المدروسة، والقيمة 0 عند عدم ظهورها. ويعرف هذا المعامل بالعلاقة التالية:

$$R_{PB} = \sqrt{\frac{n_1 \cdot n_0}{n}} \cdot \left(\frac{\bar{x}_1 - \bar{x}_0}{\sqrt{\sum \left(x - \bar{x}\right)^2}} \right) \qquad (22\text{-}10)$$

حيث إن n هو حجم العينة، و n_1 عدد حالات ظهور الصفة المدروسة، و n_0 عدد حالات عدم ظهورها.

وإن $\bar{x}_1$ هو متوسط قيم X المقابلة للقيم 1 التي يأخذها المتحول الاسمي.

وإن $\bar{x}_2$ هو متوسط قيم X المقابلة للقيم 0 التي يأخذها المتحول الاسمي.

وإن $\bar{x}$ هو متوسط قيم X في إجمالي العينة ولجميع قيم X .

ويأخذ هذا المتحول قيمه في المجال $[-1 , +1]$، وكلما كانت قيمه قريبة من الواحد كان الارتباط متيناً.

مثال(10−13):

ادرس فيما إذا كانت هناك علاقة ارتباطية بين الدخل الأسبوعي للفرد وحيازته على شهادة جامعية، وذلك من خلال البيانات التالية المأخوذة من عينة مؤلفة من 15 موظفاً.

جدول(10−24): بيانات المثال

التسلسل	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
الدخل ل.س	2400	1500	3500	3000	1200	2500	1300	2100	2700	2900	1900	2100	1400	1600	1000
حيازة شهادة عليا	1	1	0	1	0	1	0	1	1	1	0	1	0	0	0

لقد رمزنا بـ 1 لحيازة الموظف على شهادة جامعية، وبـ 0 على عدم حيازته عليها.

الحل: إن هذه العلاقة هي بين متحول كمي X هو مقدار الدخل الأسبوعي للفرد، ومتحول اسمي هو الحيازة على شهادة جامعية الذي يأخذ إحدى القيمتين 1 أو 0 حسب ظهور الخاصة، أو عدم ظهورها.

ولدراسة هذه العلاقة نحسب معامل ارتباط نقطة السلسلة المزدوجة، ولذلك نقوم بحساب عناصره حيث نجد أن:

$$n = 15 \quad , \quad n_1 = 8 \quad , \quad n_0 = 7$$

ثم نحسب متوسط قيم X المقابلة للعدد (1) فقط فنجد أن:

$$\bar{x}_1 = \frac{2400 + 1500 + 300 + 2500 + 2100 + 2700 + 2900 + 2100}{8}$$

$$\bar{x}_1 = 2400 L.S.$$

ثم نقوم بحساب متوسط قيم X المقابلة للعدد (0) فقط فنجد أن:

$$\bar{x}_0 = \frac{3500 + 1200 + 1300 + 1900 + 1400 + 1600 + 1000}{7}$$

$$\bar{x}_0 = 1700 L.S.$$

ثم نقوم بحساب المتوسط العام لجميع قيم X حيث نجد أن:

$$\bar{x} = \frac{1}{15}\sum x_i = \frac{1}{15} \cdot 31100 = 2073{,}33\, L.S.$$

وبذلك نجد أن مجوع مربعات الانحرافات عن المتوسط العام يساوي:

$$\sum_{i=1}^{15}(x_i - \bar{x})^2 = \sum_{i=1}^{15}(x_i - 2073{,}33)^2 = 7809299{,}574$$

وأخيراً نقوم بحساب معامل ارتباط السلسلة المزدوجة، فنجد أن:

$$R_{PB} = \sqrt{\frac{8 \cdot 7}{15}}\left(\frac{2400 - 1700}{780929{,}57}\right) = 0{,}484$$

وهو يدل على درجة ارتباط مقبولة بين الدخل والشهادة الجامعية.

10-7: اختبارات حول الوسيط لمجتمعين أو أكثر:

يوجد عدة اختبارات حول الوسيط لمجتمعين أو أكثر من خلال عينيتين مستقلتين أو مرتبطتين . ولذلك سنستعرض أهم هذه الاختبارات كما يلي:

10-7-1: اختبار (مان وتيني Mann - Whitney) لتساوي وسيطي متحول كمي X في مجتمعين من عينتين مستقلتين:

إن هذه الاختبار مشهور في العلوم الاجتماعية والطبية، وذلك لأنه يعتبر بديلاً عن الاختبارين المعلميتين Z و t للمتحولات الكمية . ويتم استخدامه عندما تكون الشروط المفروضة على الاختبارين Z و t غير محققة . أي أنه يطبق على المتحولات الكمية التي لا تخضع للتوزيع الطبيعي وفي الحالات التي تكون فيها حجوم العينات صغيرة (أصغر من 20 عنصراً من كل مجتمع) . أي أنه لا يشترط هنا أن يكون X طبيعياً في المجتمعين ولا يبحث عن توقعه $\mu(x)$ ولا عن تباينه $\sigma^2(x)$.

ولكنه يشترط لتطبيقه شروطاً أخرى هي:

- الشروط المفروضة على اختبار Mann – Whitney :

1- أن يكون المتحول المدروس متحولاً عشوائياً مستمراً في المجتمعين.

2- أن يتم سحب عينتين عشوائيتين من المجتمعين بحجمين n_1 و n_2 وأن تكونا مستقلتين .

3- أن يكون سلم قياس مشاهداه X في المجتمعين رتبياً على الأقل.

4- أن يكون الاختلاف بين توزيعي X في المجتمعين مرتبطاً فقط بنقطة تموضع الوسيط . وليس له علاقة بنوع التوزيع بل بشكله وبتموضع قيمه حول ذلك الوسيط.

لذلك نفترض إننا نريد دراسة تغيرات سلوك متحول كمي X، **في مجتمعين محددين** P_1 و P_2 بوسيطين مجهولين M_1 و M_2 . وإننا سحبنا منهما عينتين عشوائيتين مستقلتين بحجمين n_1 و n_2 على الترتيب .

ولنرمز للمشاهدات المأخوذة من العينة الأولى بالرموز $X_1\ X_2\ X_3 \ldots X_{n1}$ ، وللمشاهدات المأخوذة من العينة الثانية بالرموز $X'_1\ X'_2\ X'_3 \ldots . X'_{n2}$.

وبناء على ذلك نضع الفرضيتين كما يلي:

فرضية العدم: إن تموضع قيم X في المجتمع الأول متشابه مع تموضعه في المجتمع الثاني.

الفرضية البديلة: إن تموضع قيم X في المجتمع الأول يختلف عن تموضعه في المجتمع الثاني.

وللتعبير عن هاتين الفرضيتين في دراسة تموضع قيم X في المجتمعين نستعين بمفهوم الوسيط، الذي يقسم قيم X إلى قسمين متساويين: قسم على يساره (أصغر منه) وقسم على يمينه (أكبر منه)، ونضع هاتين الفرضيتين بدلالة الوسيطين M_1 و M_2 على أحد الأشكال التالية (شكل واحد فقط):

$$\begin{array}{lll} H_0: M_1 = M_2 & H_1: M_1 \neq M_2 & (A) \\ H_0: M_1 \geq M_2 & H_1: M_1 < M_2 & (B) \\ H_0: M_1 \leq M_2 & H_1: M_1 > M_2 & (C) \end{array} \qquad (23-10)$$

فإذا أخذنا الشكل (A) لهاتين الفرضيتين:

$$H_0: M_1 = M_2 \qquad\qquad H_1: M_1 \neq M_2$$

فإن هذا يعني أن وسيطي المجتمعين متساويان (بعض النظر عن التوزيع الاحتمالي)، وإن عدد المشاهدات الواقعة على طرفي هذين الوسيطين في المجتمعين متساويان.

وللاستفادة من هذه الخاصة (الواردة في مضمون H_0) نقوم بإجراء العمليات التالية:

- اجراء تطبيق اختبار (مان– ويتني):

1– ندمج مشاهدات العينتين في عينة واحدة (نسميها العينة الكلية)، ونرتب قيم تلك المشاهدات ضمن تلك العينة الكلية ترتيباً تصاعدياً، ونضعها في جدول واحد (مع الحفاظ على أصلهما في العينتين).

2– نقوم بوضع رتب متسلسلة لهذه المشاهدات المدمجة من العدد /1/ حتى العدد $(N = n_1 + n_2)$ حيث أن N هو مجموع حجمي العينتين.

3– نقوم بحساب مجموع رتب المشاهدات العائدة للعينة الأولى ونرمز له بـ R_1 , ثم نقوم بحساب مجموع رتب المشاهدات العائدة للعينة الثانية ونرمز له بـ R_2. وهنا نلاحظ أن المجموع الكلي للرتب (التي تشكل متوالية حسابية: 1 2 3 4 ...N) يساوي: $R_1 + R_2 = \frac{N}{2}(N+1)$.

4– وهنا يمكننا أن نقوم بمقارنة R_1 مع R_2 . فإذا كان R_1 متقارباً من R_2 فهذا يعني أن تموضع قيم X في المجتمعين هو تموضع عشوائي ومتداخل، وهذا يدعونا إلى قبول فرضية العدم: $H_0: M_1 = M_2$ والعكس بالعكس.

ولكن إذا كان تموضع جميع قيم X من العينة الأولى على الطرف الأيسر من الوسيط، فإن ذلك يدعونا لرفض H_0، وفي هذه الحالة فإن رتب المشاهدات فيها تأخذ الأرقام المتسلسلة للمتوالية الحسابية n_1.... 3 2 1 , وبالتالي فإن مجموعها R'_1 يساوي: $R'_1 = \frac{n_1}{2}(1+n_1)$.

وإذا كانت قيم X من العينة الثانية تقع على الطرف الأيمن من الوسيط، فإن ذلك يدعونا أيضاً لرفض H_0 , وفي هذه الحالة فإن رتب المشاهدات فيها تأخذ الأرقام المتسلسلة للمتوالية الحسابية التالية: n_2 ... ... 4 3 2 1 , وبالتالي فإن مجموعها R'_2 (في هذه الحالة) يساوي:

$$R'_2 = \frac{n_2}{2}(1+n_2) \qquad (24-10)$$

5– والآن نقوم بتعريف المؤشرين U_1 و U_2 لاختبار (مان– ويتني) كما يلي:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \qquad (25-10)$$

$$U_1 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 \qquad (26-10)$$

حيث أن: R_1 هو مجموع الرتب الفعلية لمشاهدات العينة الأولى في العينة الكلية المدمجة.

وأن: R_2 هو مجموع الرتب الفعلية لمشاهدات العينة الثانية في العينة الكلية المدمجة.

6- ثم نقوم بحساب قيمة المؤشر U لـ(مان- ويتني) من أصغر العددين U_1 وU_2 أي من العلاقة :

$$U = \min(U_1, U_2) \qquad (27-10)$$

وهنا نشير إلى أنه يمكن البرهان بسهولة على أن : $U_1 + U_2 = n_1 n_2$ في كل الأحوال, كما نلاحظ أن المؤشر U يأخذ قيمه الممكنة من الصفر حتى القيمة $n_1 n_2$. فهو يساوي الصفر عند الانفصال التام لمشاهدات العينتين , ويساوي $n_1 n_2$ في حالة التوزع العشوائي التام . وأن متوسطه =$n_1 n_2/2$.

7- نقوم باتخاذ القرار حول H_0 التي من الشكل (A) كما يلي:

نبحث في جدول U عن القيمة الحرجة $U_{\frac{\alpha}{2}}$ المقابلة لنصف مستوى الدلالة (لأن الاختبار في (A) ثنائي الجانب) وللعددين n_1 وn_2 ، ثم نقارن U مع $U_{\frac{\alpha}{2}}$ الجدولية ونتخذ القرار كما يلي:

إذا كانت: $U < U_{\frac{\alpha}{2}}$ نرفض فرضية العدم H_0 ونقبل H_1 (انتبه إلى أن ذلك معاكس لما نعرفه عن الاختبارات الأخرى).

وإذا كانت: $U \geq U_{\frac{\alpha}{2}}$ نقبل فرضية العدم H_0 التي تقول أن تموضعات قيم X في المجتمعين متشابهة.

ملاحظة: إذا كانت فرضية العدم H_0 أحادية الجانب من النوع (B) فإننا نقارن قيمة U مع U_{α} . ونرفض فرضية العدم إذا كانت $U < U_{\alpha}$. والعكس بالعكس.

أما إذا كانت فرضية العدم H_0 أحادية الجانب من النوع (C) فإننا نقارن قيمة U مع $U_{1-\alpha}$. حيث $U_{1-\alpha} = n_1 n_2 - U_{\alpha}$ ونرفض H_0 إذا كانت $U < U_{1-\alpha}$.

مثال (10-14): في دراسة لتموضع معدلات تخرج طلاب الاحصاء في جامعتي دمشق وحلب، سحبنا منهما عينتين بحجمين طالباً $n_1 = 5$ و$n_2 = 4$ فوجدنا أن معدلاتهم كانت كما يلي:

جدول(10-25): بيانات المثال مع توضيح كيفية ترتيبها ومعالجاتها

رقم الطالب	المعدلات في جامعة دمشق	المعدلات في جامعة حلب	الرقم	المعدلات المرتبة ورتبها في العينة الكلية			
				معدلات دمشق	معدلات حلب	رتب دمشق	رتب حلب
1	**78**	**90**	**1**	**50**	–	**1**	–
2	**64**	**70**	**2**	–	**51**	–	**2**
3	**75**	**53**	**3**	–	**53**	–	**3**
4	**50**	**51**	**4**	**64**	–	**4**	–
5	**82**	–	**5**	–	**70**	–	**5**
			6	**75**	–	**6**	–
			7	**78**	–	**7**	–
			8	**82**	–	**8**	–
			9	–	**90**	–	**9**
				مجموع الرتب		$R_1 = 26$	$R_2 = 19$

وبعد دمج مشاهدات هاتين العينتين وترتيب قيمها تصاعدياً ثم إعطاء رتب متسلسلة لتلك القيمة ضمن العينة الكلية نحصل على العمودين الأخيرين وعلى مجموعي الرتب كما يلي:

$$R_1 = 26 \qquad R_2 = 19$$

ثم نقوم بحساب قيمتي المؤشرين U_1 وU_2 من العلاقتين (10−25) و(10−26) فنجد أن:

$$U_1 = 5 * 4 + \frac{5 * 6}{2} - 26 = 9$$

$$U_2 = 5 * 4 + \frac{4 * 5}{2} - 19 = 11$$

ثم نقوم بحساب مؤشر الاختبار U من العلاقة:

$$U = \min(9, 11) = 9$$

ولاتخاذ القرار المناسب حول H_0 والتي نفترض أنها من الشكل (A) التالي:

$$H_0: M_1 = M_2$$

$$H_1: M_1 \neq M_2 \qquad \text{(الاختبار ثنائي الجانب)}$$

نبحث عن قيمة $U_{\frac{\alpha}{2}}$ الحرجة من جداول U والمقابلة للحجمين n_1 وn_2 ولنصف مستوى الدلالة $\alpha = 0.10$ نجد أن: $U_{0.05(4,5)} = 1$

وبالمقارنة نجد أن: $U > U_{\frac{\alpha}{2}}$ لذلك نقبل فرضية العدم H_0 ونعترف بتموضع المعدلات في الجامعتين متشابهتين. (لاحظ أن أسلوب القرار يختلف عن الأساليب العادية).

10−7−1−1: اختبار (مان− وينتي) للعينات الكبيرة (n_1 , $n_2 \geq 20$):

إذا كانت كل من n_1 وn_2 أكبر من 20 عنصراً، فإننا نستفيد من نظرية النهاية المركزية . ونعتبر مؤشر الاختبار U خاضعاً تقاربياً للتوزيع الطبيعي وتوقعه وتباينه يساويان ما يلي:

$$\mu_u = \frac{n_1\, n_2}{2} \quad , \quad \sigma_U^2 = \frac{n_1\, n_2(n_1 + n_2 + 1)}{12} \qquad (28-10)$$

وبناءً على ذلك تم تعريف مؤشر مان− ويتني الطبيعي كالتالي:

$$Z = \frac{u - \mu_u}{\sigma_u} = \frac{u - \frac{n_1\, n_2}{2}}{\sqrt{\frac{n_1\, n_2(n_1 + n_2 + 1)}{12}}} \qquad (29-10)$$

حيث أن Z يخضع تقاربياً للتوزيع الطبيعي المعياري $N(0,1)$. ولاتخاذ القرار حول H_0 نقارن Z مع $Z_{1-\frac{\alpha}{2}}$ ونتخذ القرار كالعادة.

ملاحظة: إذا كانت بعض المشاهدات في العينة الكلية متساوية فإننا نعطيها رتبة واحدة مساوية لمتوسط الرتب التي يفترض أن تكون لها . وتسمى كل حالة من هذه الحالات بالحالات العقدية . ونرمز لعددها في العينة الكلية بـ K .

فإذا كان عدد الحالات العقدية K كبيراً . فإن ذلك سيؤثر على حساب المؤشر U . لذلك نقوم بتعديل المؤشر الطبيعي Z ليصبح على الشكل التالي:

$$Z' = \frac{U - \frac{n_1\, n_2}{2}}{\sqrt{\left(\frac{n_1\, n_2}{N(N-1)}\right)\left(\frac{N^3 - N}{12} - \sum_{i=1}^{k} T_i\right)}} \qquad (30-10)$$

حيث أن: $N = n_1 + n_2$، وأن كل T_i تساوي:

$$T_i = \left(\frac{t^3 - t}{12}\right)_i \qquad (31-10)$$

وحيث أن: t هو عدد المشاهدات المتساوية في العقدة المرتبة i وأن $\sum_{i=1}^{k} T_i$ هو مجموع T_i على جميع العقد الموجودة في العينة الكلية.

10-7-1-2: اختبار ويلكوكسن لمجموع الرتب Wilcoxon For sum Ranks ($n_1, n_2 \geq 10$):

وهو اختبار شبه باختبار (مان- وينتي) وبديل له، ولكنه يعتمد في تعريفه على مجموع رتب العينة الأولى R_1 فقط، ويتطلب تطبيقه نفس شروط (مان وينتي) ويعرف بالعلاقة التالية:

$$Z = \frac{R_1 - \frac{n_1\,(N+1)}{2}}{\sqrt{\frac{n_1\, n_2(n_1 + n_2 + 1)}{12}}} \quad : \; n_1 \geq 10 \text{ حيث أن} \qquad (32-10)$$

حيث Z يخضع للتوزيع الطبيعي المعياري $N(0,1)$. وبعد حساب قيمة Z نقارنها مع $Z_{1-\frac{\propto}{2}}$ (للثنائي) ونرفض H_0 إذا كان $Z > Z_{1-\frac{\propto}{2}}$, أو نقارنها مع $Z_{1-\propto}$ (للأحادي) ونرفض H_0 إذا كان $Z > Z_{1-\propto}$.

مثال (10-15): في دراسة لدليل البدانة BMI (Body Mass Index) بين الرجال والنساء . سحبنا عينتين منهما بحجمين $n_1 = 13$ و $n_2 = 12$, وبعد إجراء القياسات وحساب قيم ذلك الدليل من العلاقة المعروفة:

$$BMI = \frac{m(kg)}{h^2(m)} = \frac{Mass(kg)}{(height)^2(\text{متر})} = \frac{\text{الوزن}(\text{كغ})}{\left[\text{الطول}(\text{متر})\right]^2}$$

حصلنا على القيم العددية لـ BMI لعناصر كلتا العينتين وضعناها في الجدول التالي:

جدول(10-26): بيانات المثال

	للنساء			للرجال		الرقم
	الرتب	قيم الدليل		الرتب	قيم الدليل	
ملاحظة: نلاحظ وجود بعض القيم	(2,5)	19,6		(11,5)	23,8	1
المتساوية ضمن القيم الموجودة في	(11,5)	23,8		(9)	23,2	2
العينة الكلية . لذلك تم إعطائها رتب	(2,5)	19,6		(14)	24,6	3
متسلسلة ثم حساب متوسط تلك	(22)	29,1		(17)	26,2	4
الرتب ووضعها أمام كل منهما مثل:	(15,5)	25,2		(10)	23,5	5

<table>
<tr><th rowspan="2">الرقم</th><th colspan="2">للرجال</th><th rowspan="2"></th><th colspan="2">للنساء</th><th rowspan="2"></th></tr>
<tr><th>قيم الدليل</th><th>الرتب</th><th>قيم الدليل</th><th>الرتب</th></tr>
<tr><td>6</td><td>24,5</td><td>(13)</td><td></td><td>21,4</td><td>(5)</td><td rowspan="8">مثل القيمتين: 19.6 و19.6 كان يجب أن نعطيهما: 2 3 لذلك نحسب المتوسط: $\frac{2+3}{2} = 2,5$ ونضع الرتبة 2.5 أمام كل منهما : وكذلك الأمر بالنسبة للقيمتين 23.8 من الرجال و23.8 من النساء وكذلك الأمر بالنسبة للقيمتين: 25.2 للرجال و25.2 للنساء</td></tr>
<tr><td>7</td><td>21,5</td><td>(6)</td><td></td><td>22,0</td><td>(7)</td></tr>
<tr><td>8</td><td>31,4</td><td>(24)</td><td></td><td>27,5</td><td>(19)</td></tr>
<tr><td>9</td><td>26,4</td><td>(18)</td><td></td><td>33,5</td><td>(25)</td></tr>
<tr><td>10</td><td>22,7</td><td>(8)</td><td></td><td>20,6</td><td>(4)</td></tr>
<tr><td>11</td><td>27,8</td><td>(20)</td><td></td><td>29,9</td><td>(23)</td></tr>
<tr><td>12</td><td>28,1</td><td>(21)</td><td></td><td>17,7</td><td>(1)</td></tr>
<tr><td>13</td><td>25,2</td><td>(15,5)</td><td></td><td>——</td><td>——</td></tr>
<tr><td>مجموع</td><td>$n_1 = 13$</td><td>$R_1 = 187$</td><td></td><td>$n_2 = 12$</td><td>$R_2 = 138$</td><td></td></tr>
</table>

ثم قمنا بدمج هذه القياسات ضمن عينة واحدة (دون ترتيب)، ولكننا قمنا بإعطائها الرتب المقابلة لقيمها المتصاعدة ضمن العينة الكلية مع الأخذ بعين الاعتبار القيم المتساوية أو المتكررة . فحصلنا على الجدول السابق . ووضعنا الرتب المقابلة للمشاهدات ضمن قوسين وفي عمودين مقابلين لها.

ثم نضع الفرضتين كما يلي:

فرضية العدم H_0: إن قيم دليل البدانة عند الرجال والنساء لهما وسيطين متساويين (لهما تموضع متشابه).

الفرضية البديلة H_1: إن قيم دليل البدانة عند الرجال والنساء لهما وسيطين غير متساويين (تموضعهما غير متشابه) .

ويمكن أن نكتب ذلك رياضياً كما يلي:

$$H_0: M_1 = M_2$$

$$H_1: M_1 \neq M_2$$ (ثنائي الجانب)

والمطلوب: اختبار صحة الفرضية H_0 بمستوى دلالة $\propto = 0.05$.

ومن الجدول السابق نجد أن مجموع رتب العينة الأولى يساوي:

$$R_1 = 11.5 + 9 + 14 + \cdots + 15.5 = 187$$

وأن مجموع رتب العينة الثانية يساوي:

$$R_2 = 2.5 + 11.5 + 2.5 + \cdots + 1 = 138$$

ثم نقوم بحساب قيمتي U_1 وU_2 فنجد أن:

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 = 13 * 12 + \frac{13 * 14}{2} - 187 = 60$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2 = 13 * 12 + \frac{12 * 13}{2} - 138 = 96$$

وبذلك نجد أن قيمة المؤشر U تساوي:

$$U = \min[60\,,96] = 60$$

ولاتخاذ قرار حول الفرضيتين السابقتين، نقوم بالبحث في جدول (مان− ويتني) عن قيمة $U_{\frac{\propto}{2}}$ المقابلة لنصف مستوى الدلالة $\frac{\propto}{2}$ وللحجمين n_1 و n_2 فنجد أنها تساوي: $U_{\frac{\propto}{2}} = 41$.

وبالمقارنة نجد أن $U > U_{\frac{\propto}{2}}$ لذلك نقبل فرضية العدم H_0 (القرار معكوس) ونعتبر أن دليلا البدانة للرجال والنساء لهما وسيطين متساويين.

ولكن هذه النتيجة محفوفة بالخطر، لأنها لم تأخذ بعين الاعتبار رتب القيم المتكررة . وللتخلص من هذه المخاطر نلجأ إلى إجراء الاختبار بواسطة العلاقة الخاصة بالعينات الكبيرة (رغم أن n_1 و $n_2 < 20$) وهي العلاقة:

$$Z = \frac{U - \frac{n_1\, n_2}{2}}{\sqrt{\frac{n_1\, n_2(n_1 + \ n_2 + 1)}{12}}} = \frac{60 - \frac{13 * 12}{2}}{\sqrt{\frac{13 * 12(13 + 12 + 1)}{12}}} = -0.979$$

وبمقارنة القيمة المطلقة لـ Z المحسوبة مع $Z_{1-\frac{\propto}{2}} = Z_{0.975} = 1.96$ نجد أن : $|Z| < Z_{1-\frac{\propto}{2}}$ ، لذلك نقبل فرضية العدم H_0 التي تقول أن دليلاً البدانة عند الرجال والنساء لهما وسيطين متساويين.

ولحساب تأثير القيم المتكررة نطبق العلاقة (10−30) ولذلك نقوم بحساب قيم T_i عند كل عقدة فنجد أن:

للقيمة الأولى 19.6 المكررة مرتين : $T_i = \frac{2^3 - 2}{12} = \frac{6}{12}$

للقيمة الثانية 22.8 المكررة مرتين : $T_i = \frac{2^3 - 2}{12} = \frac{6}{12}$

للقيمة الثالثة 25.2 المكررة مرتين : $T_i = \frac{2^3 - 2}{12} = \frac{6}{12}$

وإن مجموعها يساوي:

$$\sum T_i = \frac{6}{12} + \frac{6}{12} + \frac{6}{12} = \frac{18}{12}$$

ثم نقوم بتعويض ذلك في العلاقة (10−30) فنجد أن:

$$Z' = \frac{u - \frac{n_1\, n_2}{2}}{\sqrt{\left(\frac{n_1\, n_2}{N(N-1)}\right)\left(\frac{N^3 - N}{12} - \sum T_i\right)}} =$$

$$Z' = \frac{60 - \frac{13 * 12}{2}}{\sqrt{\left(\frac{13 * 12}{25 * 24}\right)\left[\frac{25^3 - 25}{12} - \frac{18}{12}\right]}} = \frac{-18}{18.374} = -0.9796$$

وبالمقارنة نجد أن $|Z'| < Z_{1-\frac{\alpha}{2}} = 1.96$، لذلك نقبل الفرضية H_0 كما فعلنا قبل اعتبار التكرارات. وهنا نشير إلى أن تأثير القيم المتكررة لم يكن ملحوظاً لأن عددها K وعدد تكرارات كل منها t كان قليلاً.
والآن لنقم باختبار صحة الفرضية H_0 بواسطة اختبار (ويلكوكسن) لذلك نحسب قيمة Z من العلاقة (31−10) فنجد أن:

$$Z = \frac{R_1 - \frac{n_1\,(N+1)}{2}}{\sqrt{\frac{n_1\,n_2}{12}(n_1 + \,n_2 + 1)}} = \frac{187 - \frac{13(26)}{2}}{\sqrt{\frac{13 * 12}{12}(13 + 2 + 1)}} = 0.879$$

وبمقارنة قيمة Z المحسوبة مع قيمتها الجدولية $Z_{1-\frac{\alpha}{2}} = 1.96$ نجد أن $|Z| < Z_{1-\frac{\alpha}{2}}$، ولذلك نقبل فرضية العدم H_0 ونعترف بأن دليلا البدانة عند الرجال والنساء لها وسيطين متساويين.
ملاحظة: كان يمكن عند تطبيق اختبار (ويلكوكسن) باعتبار عينة النساء هي الأولى ووضع $R_1 = 138$ و$\mu_U = 156$ ، فنحصل على نفس النتيجة .

10−7−2: اختبار كروسكال− وايلز (Kruskal- Willis) ويدعى اختبار H:

يعتبر هذا الاختبار بديلاً لتحليل التباين باتجاه واحد ANOVA في الاختبارات المعلمية، ويستخدم لاختبار الفروقات بين تغيرات متحول X في عدة مجتمعات (3 أو اكثر) من خلال عدة عينات مستقلة وذات حجوم صغيرة. وهو لا يشترط أن يكون X خاضعاً للتوزيع الطبيعي كما في ANOVA , ولا يتناول تساوي متوسطات تلك المجتمعات بل وسطائها. وبكلام آخر يُستخدم لاختبار تساوي الوسطاء, أي لاختبار فيما إذا كانت تلك العينات المستقلة مسحوبة من مجتمعات لها نفس الوسيط.
وبذلك يمكننا كتابة فرضيتي العدم والبديلة كما يلي:
فرضية العدم: إن العينات مسحوبة من مجتمعات لها نفس الوسيط.
الفرضية البديلة: إن العينات مسحوبة من مجتمعات لها وسطاء مختلفة.

$$H_0: M_1 = M_2 = \,M_3 \qquad (33-10)$$

$H_1: M_k \neq M_j$ من أجل زوج واحد على الأقل $(k\,,j)$

شروط تطبيقه:

1− أن يكون لدينا (3) عينات مستقلة على الأقل، وأن كل منها مسحوبة عشوائياً من مجتمعها, ونرمز لحجومها المتساوية أو المختلفة بـ.... n_1, n_2, n_3 .

2− إن تتضمن كل عينة (5) مشاهدات على الأقل (أي $n_i \geq 5$).

3− لا يشترط أن يكون المتحول المدروس X خاضعاً للتوزيع الطبيعي أو لأي توزيع آخر في المجتمعات المدروسة.

خطوات وإجراءات تطبيقيه:

1- ندمج المشاهدات المأخوذة من جميع العينات في عينة واحدة ونعتبرها العينة الكلية، ثم نرتب قيمها ترتيباً تصاعدياً من (1 حتى N) . حيث $(N = \sum n_i)$.

2- نحسب مجاميع رتب كل عينة على حدة، ونرمز لهذه المجاميع بالرموز التالية:

$$R_1 = \sum_{i=1}^{n_1} R_{1i} \ , \qquad R_2 = \sum_{i=1}^{n_2} R_{2i} \, , \qquad R_3 = \sum_{i=1}^{n_3} R_{3i} \ , ...$$

3- نقوم بتربيع هذه المجاميع الرتبية ثم نقسم تلك المربعات على حجوم العينات المقابلة لها، فنحصل على متوسطات مربعات تلك المجاميع كما يلي:

$$\frac{R_1^2}{n_1}, \qquad \frac{R_2^2}{n_2}, \qquad \frac{R_3^2}{n_3}, \dots\dots \qquad (34-10)$$

4- ثم نقوم بحساب قيمة مؤشر الاختبار (كروسكال- وايلز) والذي يرمز له بـ H من العلاقة التالية:

$$H = \frac{12}{N(N+1)}\left[\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \cdots + \frac{R_k^2}{n_k}\right] - 3(N+1) \qquad (35-10)$$

حيث K عدد العينات وn_i حجم العينة i و$N = \sum n_i$. وR_i هي المعرفة أعلاه, وهو اختبار أحادي يميني. وإن H يخضع تقاربياً لتوزيع χ^2 بـ $(k-1)$ درجة حرية.

5- لاتخاذ القرار حول H_0 نقوم بمقارنة قيمة H المحسوبة من العلاقة (10-35) مع قيمة $\chi^2_{\propto(k-1)}$ الحرجة ونتخذ القرار كما يلي:

إذا كانت $H \le \chi^2_{\propto(k-1)}$ نقبل فرضية العدم H_0 ونعترف بأن العينات مسحوبة من مجتمعات ذات وسطاء متساوية. أما إذا كانت $H > \chi^2_{\propto(k-1)}$ نرفض فرضية العدم H_0 ونقبل H_1 .

ملاحظة (1): إن الاختبار H هو مقياس لتباين مجاميع الرتب $R_1, R_2, \dots R_k$.

فإذا كانت العينات متماثلة، فإن رتب مشاهداتها تكون موزعة بانتظام أو بالتساوي بين تلك العينات وتكون قيمة H صغيرة نسبياً.

أما إذا كانت العينات مختلفة فإن رتب العناصر فيها تكون منخفضة في بعض العينات وتكون مرتفعة في بعضها الآخر، وهذا ما يؤثر على قيمة H ويجعلها كبيرة نسبياً . وهذا ما يجعلنا نرفض الفرضية H_0 .

ملاحظة (2): إذا كانت بعض قيم المشاهدات متساوية (متكررة)، فإننا نقوم بترتيبها بمتوسط رتبها المتسلسلة، ثم نجري تعديلاً على H بتقسيمه على المقدار $\left(1 - \frac{t^3-t}{N^3-N}\right)$.

مثال (10-16): نريد دراسة الفروقات بين درجات الطلاب في مقرر الرياضيات X في (3) مدارس. فسحبنا منها (3) عينات بحجوم متساوية $n_1 = n_2 = n_3 = 8$ ووضعنا النتائج في الجدول التالي:

جدول(10− 27): بيانات المثال

المدارس	المدرسة الأولى		المدرسة الثانية		المدرسة الثالثة	
البيان / رقم الطالب في عينة المدرسة	الدرجة X_i	الرتبة في العينة الكلية	الدرجة X_i	الرتبة في العينة الكلية	الدرجة X_i	الرتبة في العينة الكلية
1	25	(10)	11	(3)	21	(6)
2	22	(7.5)	29	(13)	5	(1)
3	31	(15)	34	(18)	18	(4)
4	20	(5)	33	(17)	23	(9)
5	45	(22)	47	(23)	22	(7.5)
6	26	(11)	37	(20)	8	(2)
7	30	(14)	38	(21)	32	(16)
8	36	(19)	48	(24)	28	(12)
المجموع	ــــ	$R_1 = 103.5$	ــــ	$R_2 = 139$	ــــ	$R_3 = 57.5$

ندمج درجات الطلاب في هذه العينات في عينة واحدة كلية، ثم نقوم بوضع رتب تصاعدية لها، فنحصل على الرتب المبينة ضمن قوسين في الجدول السابق . ثم نحسب مجاميعها فنجد أن:

$$R_1 = 103.5 \qquad R_2 = 139 \qquad R_3 = 57.5$$

ثم نقوم بحساب قيمة المؤشر H من العلاقة (10−35) فنجد أن:

$$H = \frac{12}{N(N+1)}\left[\frac{R_1^2}{n_1}+\frac{R_2^2}{n_2}+\frac{R_3^2}{n_3}\right] - 3(N+1) =$$

$$H = \frac{12}{24(25)}\left[\frac{(103.5)^2}{8}+\frac{(139)^2}{8}+\frac{(57.5)^2}{8}\right] - 3(25) =$$

$$H = 8.3488$$

ولاتخاذ القرار المناسب حول الفرضية H_0 نبحث في جداول χ^2 عن القيمة الحرجة $\chi^2_{\propto(k-1)}$ المقابلة لمستوى الدلالة $\propto = 0.05$ ولدرجة الحرجة $(k-1=3-1=2)$، فنجد أنها تساوي: $\chi^2_{0.05\ \ 2} = 6$ وبالمقارنة نجد أن $H > \chi^2_{0.05\ \ 2}$ ، لذلك نرفض فرضية العدم H_0 ، ونقبل الفرضية البديلة التي تقول إن هذه العينات مسحوبة من مجتمعات ذات وسطاء مختلفة. أي أن وسطاء درجات الطلاب في مقرر الرياضيات في هذه المدارس غير متساوية وتميل إلى المدرسة الثانية .

ملاحظة: يمكننا أن نأخذ بعين الاعتبار رتب المشاهدات المتكررة (المشاهدة 22 مكررة مرتين ورتبتها 7.5) . لذلك نقوم بتقسيم H على معامل التصحيح $\left(1-\frac{t^3-t}{N^3-N}\right)$.

فنجد أن:

$$H' = \frac{H}{1 - \frac{t^3 - t}{N^3 - N}} = \frac{8.3488}{1 - \frac{2^3 - 2}{(24)^3 - 24}} = 8.352$$

وبمقارنة هذه القيمة مع القيمة الحرجة $\chi^2_{0.05\ \ 2} = 6$ نجد أن $H' > \chi^2_{0.05\ \ 2}$ لذلك نرفض H_0 ونخلص لنفس النتيجة

10-7-3: اختبار ويلكوكسن للرتب المؤشرة (Wilcoxon Signed ranks):

وهو اختبار لا معلمي يختلف عن اختبار ويلكوكسن لمجموع الرتب، ويطبق على رتب بيانات متحول واحد X في عينتين مرتبطتين, أو على رتب بيانات مشكلة من أزواج متقابلة وناتجة عن تجربتين (X_1, X_2) على عناصر عينة واحدة , ويمكن تطبيقه على نتائج تجربة قبلية X_1 مع تجربة بعدية X_2 ، أو على نتائج تجربة معينة مع تجربة ضابطة , ويُستخدم لاختبار فيما إذا كان وسيط الفروقات بين نتائج الأزواج المتقابلة $d_i = (X_{1i} - X_{2i})$ معدوماً أو يساوي الصفر.

وبذلك يمكننا وضع فرضيتي الاختبار على الشكل التالي:

فرضية العدم: إن وسيط فروقات الأزواج المتقابلة $d_i = (x_{1i} - x_{2i})$ في المجتمع المسحوبة منه العينة معدوم أو يساوي الصفر .

الفرضية البديلة: إن وسيط فروقات الزواج المتقابلة $d_i = (x_i - y_i)$ في المجتمع المسحوبة منه العينة غير معدوم (لا يساوي الصفر).

كما يمكننا تطبيق هذا الاختبار لاختيار حول فيما إذا كان وسيط تلك الفروقات في المجتمع يساوي قيمة معينة . وعندها يتم تعديل فرضية العدم بما يتناسب مع ذلك.

- **متطلبات تطبيق الاختبار:**

1- أن تكون البيانات مؤلفة من أزواج متقابلة لقيم تجربتين X_1 و X_2 مأخوذة من عينة واحدة مسحوبة عشوائياً من المجتمع.

2- أن يكون لجملة الفروقات توزيع متناظر تقاربياً حول قيمة الوسيط . دون أن يشترط أن تخضع تلك البيانات للتوزيع الطبيعي.

- **لصياغة مؤشر الاختبار نقوم بالإجراءات التالية:**

1- نقوم بحساب الفروقات بين عنصري الأزواج المتقابلة d_i، وهي قد تكون موجبة أو سالبة أو معدومة . ونحتفظ بإشاراتها . ولكننا نهمل الأزواج التي يكون فيها $d_i = 0$، ولنفرض أن عدم الفروقات غير المعدومة يساوي K .

2- نتجاهل مؤقتاً إشارات الفروقات d_i (غير المعدومة) . ونقوم بترتيب تلك الفروقات (حسب قيمها المطلقة) تصاعدياً وإعطائها أرقاماً متسلسلة من (1 حتى K)، (حيث K عدد الفروقات غير المعدومة), وعندما يكون بعض قيم d_i متساوية نعطيها جميعاً رقماً يساوي متوسط رتبها المتسلسلة، كما فعلنا في الاختبارات السابقة.

3- نعود ونضع أمام كل من أرقام الرتب (المذكورة في الخطوة 2) إشارة الفرق d_i المقابل لها. وهذا يعني إدخال الإشارات التي تجاهلناها في الخطوة (2) على الرتب، فنحصل على متوالية متناوبة من الرتب مثل: ... $+r_1 - r_2 - r_3 + r_4$.

4- نقوم بإيجاد مجموع القيم المطلقة للرتب السالبة ونرمز له بـ T^- . وكذلك نقوم بإيجاد مجموع قيم الرتب الموجبة ونرمز له بـ T^+ ، فيكون لدينا:

$T^- = \sum_{i=1}^{k_1} |r_i < 0|$ و $T^+ = \sum_{i=1}^{k_2} |r_i > 0|$.

5- لنأخذ أصغر المجموعين T^- و T^+ ، ولنرمز له بـ T فيكون لدينا من الخطوة (4) ما يلي:

$$T = \min[T^- , T^+] \qquad (36-10)$$

6- لنفترض أن K هو عدد الأزواج (x_i , y_i) التي فروقاتها d_i غير معدومة، ونضع فرضيتي الاختيار كما يلي:

$$H_0 : M_{X_1} = M_{X_2} \qquad H_1 : M_{X_1} \neq M_{X_2}$$

7- نقوم بحساب قيمة مؤشر الاختبار المعرف من إحدى العلاقتين التاليتين (حسب العدد K):

أ- إذا كان عدد الفروقات d_i غير المعدومة $K \leq 30$ ، فإننا نعتبر المجموع الأصغر T هو مؤشر الاختبار المناسب . ولاتخاذ القرار حول H_0 نقارن T مع القيمة الحرجة $T_{\frac{\propto}{2}}$ ونتخذ القرار كما يلي (المتراجحة معكوسة كما في مان- ويتني).

إذا كان $T < T_{\frac{\propto}{2}}$ نرفض فرضية العدم H_0 ونقبل H_1 .

وإذا كان $T \geq T_{\frac{\propto}{2}}$ نقبل فرضية العدم H_0 .

ب- إذا كانت $K > 30$ فإننا نستخدم مؤشر اختبار (ويلكوكسن) المعياري Z المعرف بالعلاقة:

$$Z = \frac{T - \frac{K(K+1)}{4}}{\sqrt{\frac{K(K+1)(2K+1)}{24}}} \qquad (37-10)$$

ج- إذا كانت العينة تتضمن مشاهدات متساوية ومكررة t مرة فإننا نعدل مقام العلاقة (10-37) ونطرح منه $\frac{\sum(t3-t)}{48}$ ونتابع الاختبار.

ولاتخاذ القرار حول H_0 في هذه الحالة نقارن قيمة Z المحسوبة مع قيمتها الحرجة $Z_{1-\frac{\propto}{2}}$ ونتخذ القرار كالعادة كما يلي:

إذا كانت $|Z| < Z_{1-\frac{\propto}{2}}$ نقبل فرضية العدم H_0 .

أما إذا كانت $|Z| \geq Z_{1-\frac{\propto}{2}}$ نرفض فرضية العدم H_0 ونقبل الفرضية H_1 .

مثال (10−17): في دراسة على عينة بحجم $n = 8$ طلاب لدراسة الفروقات بين تحصيلهم العلمي في مقرري الرياضيات والإحصاء وجدنا أن نتائجهم كانت كما يلي:

جدول(10−28): بيانات المثال

المقررات / عناصر العينة	درجة الرياضيات	درجة الاحصاء	الفروقات	الرتب المؤشرة
1	82	63	19	7
2	69	42	27	8
3	73	74	−1	−1
4	43	37	6	4
5	58	51	7	5
6	56	43	13	6
7	76	80	−4	−3
8	65	82	3	2

الحل: نضع الفرضيتين كما يلي: $H_1 : M_{X_1} \neq M_{X_2}$ $\qquad H_0 : M_{X_1} = M_{X_2}$

وبعد حساب قيم الفروقات ثم ترتيبها تصاعدياً حسب قيمها المطلقة وإعطائها رتباً متسلسلة ثم إضافة الإشارات إلى هذه الرتب نحصل على العمودين الأخيرين في الجدول السابق، ومنه نجد أن مجموع الرتب ذات الإشارات الموجبة يساوي:

$$T^+ = 7 + 8 + 4 + 5 + 6 + 2 = 32$$

وأن مجموع الرتب ذات الإشارات السالبة (بالقيمة المطلقة).

$$T^- = 1 + 3 = 4$$

وبذلك نجد أن أصغرهما T يساوي:

$$T = \min[4\,, 32] = 4$$

وبما أن عدد الفروقات غير المعدومة $K = 8$ وأن $K < 30$ (K يساوي حجم العينة n لعدم وجود فروقات معدومة) .

فإنه لاتخاذ قرار حول الفرضية H_0، نقوم بمقارنة T مع القيمة الحرجة لها $T_{\frac{\propto}{2}}$ المأخوذة من توزيع T ، حيث نجد أن عندما $\propto = 0.05$ فإن: $T_{\frac{\propto}{2}} = 4$ ، وبالمقارنة نجد أن: $T \geq T_{\frac{\propto}{2}} = 4$, لذلك نقبل فرضية العدم H_0 (رغم إنها واقعة على الحد بين الرفض والقبول) .

وإذا قمنا باستخدام العلاقة (10−37) المخصصة للعينات الكبيرة (تجاوزاً) لإجراء هذا الاختبار نجد أن:

$$Z = \frac{4 - \frac{8*9}{4}}{\sqrt{\frac{8*9*17}{24}}} = -1.96$$

ولإجراء الاختبار نقارن القيمة المطلقة لـ Z المحسوبة مع القيمة الحرجة $Z_{1-\frac{\propto}{2}}$, والتي تساوي $Z_{0.975} =$ 1.96، وبالمقارنة نجد أن $|Z| \leq Z_{1-\frac{\propto}{2}} = 1.96$, لذلك نقبل فرضية العدم H_0 (رغم أن تمثل حالة حدية).

ولإزالة هذا الالتباس نقوم بزيادة حجم العينة إلى $n = k = 10$, ونفترض أن عدد الفروقات السالبة لم يتغير ، فنجد أن:

$$Z = \frac{4 - \frac{10 * 11}{4}}{\sqrt{\frac{10 * 11 * 21}{24}}} = -2.3955$$

وبذلك نحصل على أن: $|Z| > Z_{\frac{\propto}{2}}$ ، وعندها نرفض فرضية العدم H_0 بكل ثقة.

10-7-4: اختبار فريدمان (Friedman):

يعتبر هذا الاختبار اللامعلمي بديلاً عن الاختبار المعلمي تحليل التباين باتجاهين (two-way ANOVA) ويُطبق عندما تكون شروط ANOVA غير محققة وتكون حجوم العينات صغيرة .

ويُستخدم هذا الاختبار لتحليل التباين باتجاهين بواسطة رتب المتحول المدروس X ضمن كل عينة .

شروط وإجرءات تطبيقه:

1- أن يكون المتحول التابع المدروس X كمياً ومستمراً.

2- أن يكون لدينا b مجتمعاً $(b \geq 3)$، نسحب منها b عينة عشوائية أو نأخذ منها b نموذجاً عشوائياً ، ويطلق على هذه العينات اسم (بلوكات BLOCKS) ونرمز لحجومها بـ n_i .

3- أن نقوم بإجراء تجارب أو معالجات مستقلة على كل من هذه العينات (أي على كل من BLOCKS)، ويطلق على هذه التجارب اسم المعالجات (treatments) ,ولنرمز لعددها بـ k .

4-أن لا يكون أي تداخل بين العينات (BLOCKS) والمعالجات (treatments).

5-أن تكون قيم المشاهدات ضمن كل عينة (BLOCK) قابلة للترتيب وفق نظام معين ونرمز لمجموع رتبها في المعالجة j بالرمز R_j .

6-نسجل نتائج تلك المعالجات على تلك العينات . ولنرمز لقيم المتحول المدروس X بالرمز x_{ij} (نتيجة المعالجة j في العينة i)، ونضعها في جدول مناسب كالجدول (10-25) التالي:

الجدول (10−29): قيم المشاهدات لنتائج التجارب على العينات مع رتبها r_{ij}

المعالجات treatments / العينات (BLOCKS)	1	2	3	…	k	جدول الرتب حسب المعالجات وضمن كل عينة			
1	x_{12}	x_{12}	x_{13}	…	x_{1k}	r_{12}	r_{12}	…	r_{1k}
2	x_{21}	x_{22}	x_{23}	…	x_{2k}	r_{21}	r_{22}	…	r_{2k}
3	x_{31}	x_{32}	x_{33}	…	x_{3k}	r_{31}	r_{32}	…	r_{3k}
⋮	⋮	⋮	⋮	…	⋮	⋮	⋮	…	⋮
B	x_{b1}	x_{b2}	x_{b3}	…	x_{bk}	r_{b1}	r_{b2}	…	r_{bk}
مجموع الرتب عمودياً بالنسبة لكل معالجة						R_1	R_2	…	R_k

1− نقوم بوضع الفرضيتين على الشكل التالي:

فرضية العدم H_0: إن قيم الوسيط لتلك المعالجات متساوية .

$$H_0 : M_1 = M_2 = M_3 = \cdots = M_K \qquad (38-10)$$

فرضية البديلة H_1: إن قيم الوسيط لتلك المعالجات غير متساوية .

$$H_1 : M_i \neq M_h \qquad : \qquad (i\,,h) \text{ من أجل زوج واحد على الأقل}$$

ويمكن وضع الفرضيتين على الشكل التالي:

فرضية العدم H_0: إن المعالجات لها نفس التأثير على تلك المجتمعات.

فرضية البديلة H_1: إن اثر المعالجات يختلف من مجتمع لآخر.

2− نقوم بحساب مؤشر الاختبار (فريدمان) المعروف بالرمز χ_F^2 بالعلاقة التالية:

$$\chi_F^2 = \frac{12}{b * k(k+1)} \sum R_i^2 - 3b(k+1) \qquad (39-10)$$

وهو يخضع لتوزيع χ^2 بـ $(k-1)$ درجة حرية.

وهناك تعريف آخر لمؤشر اختيار (فريدمان) ويرمز له بالرمز W ويعطى من خلال العلاقة المزدوجة التالية:

$$W = \frac{12[\sum R_i^2] - 3b^2k(k+1)^2}{b^2k(k^2-1)} = \frac{\chi_F^2}{b(k-1)} \qquad (40-10)$$

وعندما يكون لدينا مشاهدات متساوية (عقدة أو أكثر) نقوم بأخذها بعين الاعتبار ونقوم بتعديل المقام في W ونجعله مساوياً لما يلي:

$$b^2k(k^2-1) - b\left(\sum t^3 - \sum t\right) \qquad (41-10)$$

حيث t هو عدد المشاهدات في العقدة المفروضة وفي أي عينة أو بلوك.

3− لاتخاذ القرار حول H_0 بواسطة العلاقة (10− 39)، نقوم بمقارنة χ^2 , فإذا كانت $\chi^2 < \chi_{\propto}^2$ نقبل H_0 والعكس بالعكس. ولكن إذا استخدمنا العلاقة (10− 40) فإننا نقوم بمقارنة القيمة المحسوبة W مع القيمة الحرجة $W_{(1-\propto)}$ والمقابله لدرجتي الحرجة $(b\,,k)$، فإذا كانت $W < W_{(1-\propto)}$ نقبل فرضية العدم H_0 والعكس بالعكس.

مثال (10−18): لنفترض أن طلاب السنة الأولى موزعين على (3) شعب وتريد إدارة الكلية دراسة الفروقات بين هذه الشعب من حيث التحصيل العلمي . فقامت بسحب عينة عشوائية من طلاب كل شعبة بحجوم $n_1 = 15$, $n_2 = 10$, $n_3 = 20$ وتابعت نتائجهم في أربعة مقررات وحسبت متوسطات علاماتهم في كل مقرر فكانت النتائج كما يلي:

جدول (10−30): متوسطات علامات طلاب كل عينة في المقررات الأربعة مع رتبها ضمن كل عينة.

المقررات أو المعالجات العينات	متوسط الرياضيات	متوسط الاحصاء	متوسط المحاسبة	متوسط الادارة	المجموع
1 $(n_1 = 15)$	70 (2)	80 (3)	90 (4)	60 (1)	300
2 $(n_2 = 10)$	65 (1)	85 (3)	75 (2)	95 (4)	320
3 $(n_3 = 20)$	90 (4)	65 (1)	80 (3)	70 (2)	305
مجموع الرتب	$R_1 = 7$	$R_2 = 7$	$R_3 = 9$	$R_4 = 7$	

الحل: نضع الفرضيتين كما يلي:

$$H_0 : M_1 = M_2 = M_3$$

$H_1 : M_i \neq M_j$: من أجل زوج واحد على الأقل

وبعد التحقق من توفر الشروط المطلوبة، نقوم بترتيب قيم متوسطات المشاهدات ضمن كل عينة برتب متصاعدة من (1) حتى (4)، فنحصل على الرتب المكتوبة ضمن قوسين وعلى مجاميعها العمودية في كل مقرر، كما هو موضح في الجدول السابق.

ثم نقوم بحساب قيمة المؤشر χ_F^2 من العلاقة (10−39) فنجد أن:

$$\chi_F^2 = \frac{12[7^2 + 7^2 + 9^2 + 7^2]}{3 * 4 * 5} - 3 * 3 * 5 = 0.6$$

ولاختبار الفرضية H_0 نبحث عن القيمة الحرجة $\chi_{\propto}^2$ المقابلة لـ $(k-1)$ درجة حرية ولنصف مستوى الدلالة $\propto$ فنجد أن: $\chi_{0.025\ 3}^2 = 7.815$, وبالمقارنة نجد أن $\chi^2 < \chi_{\frac{\propto}{2}}^2$, لذلك نقبل فرضية العدم H_0 التي تقول أن وسطاء علامات الطلاب في هذه الشعب متساوية .

مثال (10−19): في دراسة لتأثير كمية الكوبالت (CO) على صلابة الفولاذ، قام الباحثون بتحضير (8) نماذج من الحديد وطبقوا عليها، أي على قطعة من كل منها، (4) طرائق من المعالجة بواسطة الكوبالت (كنسب مئوية) فكانت درجات صلابة الفولاذ كما يلي :

جدول (10−31): نتائج المعالجة من المصدر (*Wyane w. p. 262*)

المعالجات (%co) النماذج(Block s)	A	B	C	D	A	B	C	D
					رتب المشاهدات ضمن كل نموذج			
1	44.3	45.8	45.5	44.4	1	4	3	2
2	48.3	48.7	46.9	48.8	2	3	1	4
3	49.8	48.7	56.0	48.6	3	2	4	1
4	49.8	51.3	55.3	58.6	1	2	3	4
5	56.6	56.1	58.6	54.6	3	2	4	1
6	57.6	57.5	58.1	57.7	2	1	4	3
7	72.0	74.2	89.6	83.1	1	2	4	3
8	88.1	88.7	92.6	88.2	1	3	4	2
	مجموع الرتب (عمودياً)				14	19	27	20

وبعد حساب رتب المشاهدات ضمن كل سطر (ضمن كل نموذج) ووضعها في نفس الجدول السابق وحساب مجاميعها العمودية، نقوم بحساب قيمة مؤشر الاختيار W من العلاقة (10−40)، فنجد أن:

$$\mathrm{W}=\frac{12[14^2+19^2+27^2+20^2]-3(8^2)(4)(4+1)^2}{(8)^2(4)(4^2-1)}=0.26875$$

ومن جدول قيم W نجد أن احتمال الدلالة P المقابل لها ولـ b = 8 و $k=4$ يساوي $P=0.911$ وهو أكبر من مستوى الدلالة 0.05 =∝ . لذلك نقبل فرضية العدم H_0، التي تقول أن تأثير المعالجات بالكوبات على النماذج الحديدية المذكورة ليس معنوياً وباحتمال ثقة لا يقل عن *0.95*

ملاحظة: كان يمكن استخدام المؤشر χ_r^2 . ونحسبه من العلاقة:

$$\chi^2=W*b(k-1)=0.26875*8*3=6.45$$

وبمقارنة هذه القيمة المحسوبة $\chi^2=6.45$ مع القيمة الحرجة $\chi^2_{1-\propto}$ المقابلة لمستوى الدلالة 0.05 =∝ ولدرجة الحرية $k-1=3$ والمساوية لـ $\chi^2_{0.05\,,3}=7.815$، نجد أن: $\chi^2<\chi^2_{1-\propto}$ ، لذلك نقبل أيضاً فرضية العدم H_0، ونقول بأن تأثير المعالجات بالكوبالت على النماذج المذكورة ليس معنوياً .

لجدول(10−3): دليل جدولي لتحديد الاختبار المناسب لبيانات الباحث

الاختبار الاحصائي المناسب	Q_5 ماهو عدد العينات	Q_4 ماهو نوع البيانات	Q_3 هل العينات	Q_2 ماهو الهدف	Q_1
	عدد المجموعات أو العينات	نوع البيانات (الطبيعية)	مستقلة/ أم مرتبطة (أزواج)	الفروقات أم الارتباط	لمتحول أم/ لعدة متحولات
t أو $Z - Test$ Student's Test-t or Z	2	مستمرة (طبيعية)	مستقلة (غير زوجية)	اختبارات الفروقات	لمتحول واحد
One-way ANOVA	> 2				
Man- whitney Test or Wilcoxon rank sum	2	مستمرة ولكنها (غير طبيعية) رتبية			
Kruskal- willis Test	> 2				
χ^2- *Test/ Fisher exact*	2	اسمية			
χ^2- *Test*	$> n$				
Paried t- Test	2	مستمرة (طبيعية)	مرتبطة (كأزواج)		
Repeated ANOVA	> 2				
Wilcoxon signed Ranks Test	2	مستمرة ولكنها (غير طبيعية) رتبية			
Friedman Test	> 2				
Mcnemar's Test	2	اسمية			
Pearson's correlation (r)	−	مستمرة (طبيعية)	اختبارات الارتباط لمتحولين		
Spearman's correlation (p)	−	مستمرة (وغير طبيعية) رتبية			
Assocaition/ kappa	2	اسمية (ولها حالتان فقط)			
Linear Regression		مستمرة			لعدة متحولات (متعدد)
Ordered Logistic Reg		رتبية			
Binary Logistic Reg	(حالتان)	اسمية			
Multinemial Logistic Reg	اكثر من حالتين	اسمية			

ملاحظة: لاستخدام هذا الجدول على الباحث أن يجيب على الأسئلة التالية:

Q_1 – ما عدد المتحولات: متحول واحد أم عدة متحولات,

Q_2 – ما هو هدف الاختبار : دراسة الفروقات أم دراسة الارتباط

Q_3 – ما هي صفة العينات: مستقلة مرتبطة

Q_4 – ما هو نوع البيانات: مستمرة وطبيعية مستمرة وغير طبيعية اسمية

Q_5 – ما هو عدد المجموعات: 2 اكثر من 2

ثم يقوم بتحديد الاختبار المناسب له وتطبيقه .

تمرينات

1 . في دراسة لقوة العمل حسب الحالة العملية والجنس تبين لنا ما يلي:

الحالة العملية / الجنس	مشتغل	متعطل
ذكور	320	150
إناث	130	80

فهل هناك اقتران بين الجنس والحالة العملية؟ (اختبر بمستوى دلالة $\alpha = 0{,}05$).

2 . إذا كان توزع السكان لبلدة معينة حسب الحالة التعليمية والجنس كما يلي:

الحالة العملية / الجنس	أمي	ملم	ابتدائي	إعدادي	ثانوي	جامعي	عليا
ذكور	100	130	80	150	200	50	10
إناث	180	150	90	10	70	20	5

والمطلوب:

. دراسة وجود ارتباط أو توافق بين الحالة التعليمية والجنس وذلك بمستوى دلالة $\alpha = 0{,}05$.

. اختبار فيما إذا كان التوزيع التجريبي للذكور يخضع للتوزيع المنتظم، وذلك بمستوى دلالة $\alpha = 0{,}10$.

. اختبار فيما إذا كان التوزيع التجريبي للإناث يخضع للتوزيع المنتظم، وذلك بمستوى دلالة $\alpha = 0{,}10$.

. اختبار فيما إذا كان التوزيع التجريبي للذكور يخضع لتوزيع المجموع (ذكور + إناث)، وذلك بمستوى دلالة $\alpha = 0{,}05$.

3 . إذا كان توزع النساء حسب أعمارهن عند الزواج الأول كما يلي:

العمر عند الزواج الأول	15	16	17	18	19	20	21	22	23	24	25	25+
التكرارات	10	15	20	30	50	60	80	100	70	50	30	20

فهل تدل هذه النتائج على أن أعمار النساء عند الزواج تخضع للتوزيع الطبيعي العام؟ وما هو متوسط تباينه؟ اختبر بمستوى دلالة $\alpha = 0{,}05$.

4 . لنفترض أن تبويب 8عمال حسب الحالة التعليمية وعدد أفراد الأسرة أعطانا الجدول التالي:

الحالة التعليمية / **عدد أفراد الأسرة**	**أمي**	**ابتدائي**	**إعدادي**	**ثانوي**
3	30	40	50	70
4	40	50	40	80
5	50	60	30	90
6	60	80	20	60
7	70	90	10	40
8	80	70	5	20
9	100	60	8	10

والمطلوب:

حساب معامل ارتباط كيندال من النوع c ثم اختار معنويته بمستوى دلالة $\alpha = 0,05$.

5 . ادرس فيما إذا كان هناك علاقة بين الدخل والجنس من المعطيات التالية:

الدخل	5500	7600	4500	8000	7500	6500	5600	6800
الجنس	ذكر	ذكر	أنثى	ذكر	أنثى	أنثى	ذكر	ذكر

وذلك باستخدام كل من معامل كوريتون، ثم معامل ارتباط نقطة السلسلة المزدوجة.

مراجع الجزء الثاني

المراجع باللغة الأجنبية:

1- Anderson, T. W. (2002). An Introduction to Multivariate Statistical Analysis. 2th ed. John Wiley& sons, New York .
2- Cherstmann, S. Van, A. (2006). Robust Estimation of Cronbache's Alpha. TMA 97 P. 1660 .
3- Johnson, R. Wichern, D. (1988). Applied Multivariate Stastistical Analysis. 2th ed. Pre Hall .
4- Gopal, K. Kauji, (2006). 100 Statistical Tests. 3th ed. SG Pwal, London .
5- Kendall, M. Staley, A. (1978). The Advanced Theory of Statistics. London ch. Co. (Russian) .
6- Metropolsky, A. K. (1982). Technica of Statistical Colculetions. Moscow Nawoka .
7- Troil, M. Troil, F. (2006). Biostatistics For Biological and Health Sciences. New York, London .
8- Wayne, W. (1990), Applied Nonparametric Statistics. 2th ed PWS. Kent Publishing Company, Boston .
9- Weenink, D. (2003), Canonical correlation Analysis. Institute of Phontic Science, University of Amsterdam, IFA Proceeding 25.
10- Wilks, S. (1967), Mathematical Statistics.(Russian) Hawka, Moscow .
11- www/:Wikipedia + www/:Research Gate .

المراجع باللغة العربية:

1- درويش، رمضان محمد. (1997)، الاختبارات الاحصائية في التربية وعلم النفس. جامعة دمشق- كلية التربية .

2- الطويل، ليلى. (2015)، منهجية البحث العلمي. جامعة تشرين- كلية الاقتصاد .

3- العشعوش، ايمن+ العربيد، عدنان. (2015)، الاقتصاد القياسي. جامعة تشرين- كلية الاقتصاد

4- العلي، ابراهيم محمد. (2002)، مبادئ علم الاحصاء. جامعة تشرين- كلية الاقتصاد .

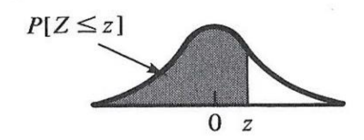
5- العلي، ابراهيم محمد+ كابوس, أمل. (1986)، الاحصاء الرياضي. جامعة حلب- كلية الاقتصاد.

6- العلي، ابراهيم محمد+ عكروش, محمد. (2005)، الاحصاء التطبيقي. جامعة تشرين- كلية الاقتصاد .

7- العلي، ابراهيم محمد. (2020)، أسس التحليل الاحصائي متعدد المتغيرات. كتاب منشور على Research Gate وعلى الموقع الخاص Dr-Alali.com وعلى صفحته الخاصة.

الجداول الاحصائية

TABLE 1 STANDARD NORMAL PROBABILITIES

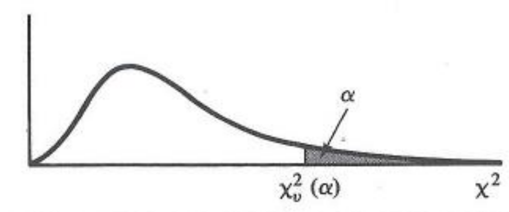


z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998

TABLE 2 STUDENT'S t-DISTRIBUTION CRITICAL POINTS

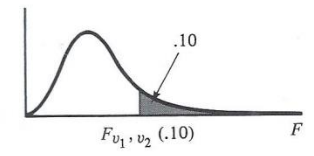
d.f. ν	α .250	.100	.050	.025	.010	.00833	.00625	.005
1	1.000	3.078	6.314	12.706	31.821	38.190	50.923	63.657
2	.816	1.886	2.920	4.303	6.965	7.649	8.860	9.925
3	.765	1.638	2.353	3.182	4.541	4.857	5.392	5.841
4	.741	1.533	2.132	2.776	3.747	3.961	4.315	4.604
5	.727	1.476	2.015	2.571	3.365	3.534	3.810	4.032
6	.718	1.440	1.943	2.447	3.143	3.287	3.521	3.707
7	.711	1.415	1.895	2.365	2.998	3.128	3.335	3.499
8	.706	1.397	1.860	2.306	2.896	3.016	3.206	3.355
9	.703	1.383	1.833	2.262	2.821	2.933	3.111	3.250
10	.700	1.372	1.812	2.228	2.764	2.870	3.038	3.169
11	.697	1.363	1.796	2.201	2.718	2.820	2.981	3.106
12	.695	1.356	1.782	2.179	2.681	2.779	2.934	3.055
13	.694	1.350	1.771	2.160	2.650	2.746	2.896	3.012
14	.692	1.345	1.761	2.145	2.624	2.718	2.864	2.977
15	.691	1.341	1.753	2.131	2.602	2.694	2.837	2.947
16	.690	1.337	1.746	2.120	2.583	2.673	2.813	2.921
17	.689	1.333	1.740	2.110	2.567	2.655	2.793	2.898
18	.688	1.330	1.734	2.101	2.552	2.639	2.775	2.878
19	.688	1.328	1.729	2.093	2.539	2.625	2.759	2.861
20	.687	1.325	1.725	2.086	2.528	2.613	2.744	2.845
21	.686	1.323	1.721	2.080	2.518	2.601	2.732	2.831
22	.686	1.321	1.717	2.074	2.508	2.591	2.720	2.819
23	.685	1.319	1.714	2.069	2.500	2.582	2.710	2.807
24	.685	1.318	1.711	2.064	2.492	2.574	2.700	2.797
25	.684	1.316	1.708	2.060	2.485	2.566	2.692	2.787
26	.684	1.315	1.706	2.056	2.479	2.559	2.684	2.779
27	.684	1.314	1.703	2.052	2.473	2.552	2.676	2.771
28	.683	1.313	1.701	2.048	2.467	2.546	2.669	2.763
29	.683	1.311	1.699	2.045	2.462	2.541	2.663	2.756
30	.683	1.310	1.697	2.042	2.457	2.536	2.657	2.750
40	.681	1.303	1.684	2.021	2.423	2.499	2.616	2.704
60	.679	1.296	1.671	2.000	2.390	2.463	2.575	2.660
120	.677	1.289	1.658	1.980	2.358	2.428	2.536	2.617
∞	.674	1.282	1.645	1.960	2.326	2.394	2.498	2.576

TABLE 3 χ^2 CRITICAL POINTS



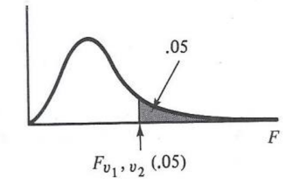
d.f. ν	.990	.950	.900	α .500	.100	.050	.025	.010	.005
1	.0002	.004	.02	.45	2.71	3.84	5.02	6.63	7.88
2	.02	.10	.21	1.39	4.61	5.99	7.38	9.21	10.60
3	.11	.35	.58	2.37	6.25	7.81	9.35	11.34	12.84
4	.30	.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	.55	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	.87	1.64	2.20	5.35	10.64	12.59	14.45	16.81	18.55
7	1.24	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.65	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.95
9	2.09	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.56	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	3.05	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.57	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	4.11	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.66	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	5.23	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.81	7.96	9.31	15.34	23.54	26.30	28.85	32.00	34.27
17	6.41	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	7.01	9.39	10.86	17.34	25.99	28.87	31.53	34.81	37.16
19	7.63	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58
20	8.26	10.85	12.44	19.34	28.41	31.41	34.17	37.57	40.00
21	8.90	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40
22	9.54	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80
23	10.20	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18
24	10.86	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.56
25	11.52	14.61	16.47	24.34	34.38	37.65	40.65	44.31	46.93
26	12.20	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29
27	12.88	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.64
28	13.56	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99
29	14.26	17.71	19.77	28.34	39.09	42.56	45.72	49.59	52.34
30	14.95	18.49	20.60	29.34	40.26	43.77	46.98	50.89	53.67
40	22.16	26.51	29.05	39.34	51.81	55.76	59.34	63.69	66.77
50	29.71	34.76	37.69	49.33	63.17	67.50	71.42	76.15	79.49
60	37.48	43.19	46.46	59.33	74.40	79.08	83.30	88.38	91.95
70	45.44	51.74	55.33	69.33	85.53	90.53	95.02	100.43	104.21
80	53.54	60.39	64.28	79.33	96.58	101.88	106.63	112.33	116.32
90	61.75	69.13	73.29	89.33	107.57	113.15	118.14	124.12	128.30
100	70.06	77.93	82.36	99.33	118.50	124.34	129.56	135.81	140.17

TABLE 4 F-DISTRIBUTION CRITICAL POINTS ($\alpha = .10$)



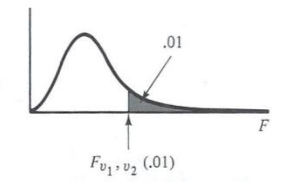
v_2 \ v_1	1	2	3	4	5	6	7	8	9	10	12	15	20	25	30	40	60
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.05	62.26	62.53	62.79
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.17	5.17	5.16	5.15
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.81	2.80	2.78	2.76
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.57	2.56	2.54	2.51
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.27	2.25	2.23	2.21
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.17	2.16	2.13	2.11
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.03	2.01	1.99	1.96
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.93	1.91	1.89	1.86
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.89	1.87	1.85	1.82
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.86	1.84	1.81	1.78
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.83	1.81	1.78	1.75
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.80	1.78	1.75	1.72
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.78	1.76	1.73	1.70
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.76	1.74	1.71	1.68
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.74	1.72	1.69	1.66
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.71	1.69	1.66	1.62
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.68	1.66	1.63	1.59
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.67	1.65	1.61	1.58
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.66	1.64	1.60	1.57
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.65	1.63	1.59	1.56
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.64	1.62	1.58	1.55
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.63	1.61	1.57	1.54
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.50	1.48	1.44	1.40
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24

TABLE 5 F-DISTRIBUTION CRITICAL POINTS (α = .05)



v_2 \ v_1	1	2	3	4	5	6	7	8	9	10	12	15	20	25	30	40	60
1	161.5	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	246.0	248.0	249.3	250.1	251.1	252.2
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.46	19.46	19.47	19.48
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.63	8.62	8.59	8.57
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.52	4.50	4.46	4.43
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.83	3.81	3.77	3.74
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.40	3.38	3.34	3.30
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.11	3.08	3.04	3.01
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.89	2.86	2.83	2.79
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.73	2.70	2.66	2.62
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.60	2.57	2.53	2.49
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.50	2.47	2.43	2.38
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.41	2.38	2.34	2.30
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.34	2.31	2.27	2.22
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.28	2.25	2.20	2.16
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.23	2.19	2.15	2.11
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.18	2.15	2.10	2.06
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.14	2.11	2.06	2.02
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.07	2.04	1.99	1.95
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.02	1.98	1.94	1.89
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.00	1.96	1.91	1.86
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.97	1.94	1.89	1.84
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.94	1.90	1.85	1.80
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.92	1.88	1.84	1.79
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.89	1.85	1.81	1.75
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.88	1.84	1.79	1.74
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.78	1.74	1.69	1.64
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.69	1.65	1.59	1.53
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.60	1.55	1.50	1.43
∞	3.84	3.00	2.61	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.51	1.46	1.39	1.32

TABLE 6 *F*-DISTRIBUTION CRITICAL POINTS ($\alpha = .01$)



ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	25	30	40	60
1	4052.	5000.	5403.	5625.	5764.	5859.	5928.	5981.	6023.	6056.	6106.	6157.	6209.	6240.	6261.	6287.	6313.
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.58	26.50	26.41	26.32
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.91	13.84	13.75	13.65
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.45	9.38	9.29	9.20
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.30	7.23	7.14	7.06
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.06	5.99	5.91	5.82
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.26	5.20	5.12	5.03
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.71	4.65	4.57	4.48
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.31	4.25	4.17	4.08
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.01	3.94	3.86	3.78
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.76	3.70	3.62	3.54
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.57	3.51	3.43	3.34
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.41	3.35	3.27	3.18
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.28	3.21	3.13	3.05
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.16	3.10	3.02	2.93
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.07	3.00	2.92	2.83
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	2.98	2.92	2.84	2.75
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.91	2.84	2.76	2.67
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.84	2.78	2.69	2.61
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.79	2.72	2.64	2.55
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.73	2.67	2.58	2.50
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.69	2.62	2.54	2.45
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.64	2.58	2.49	2.40
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.60	2.54	2.45	2.36
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.57	2.50	2.42	2.33
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.54	2.47	2.38	2.29
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.51	2.44	2.35	2.26
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.48	2.41	2.33	2.23
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.45	2.39	2.30	2.21
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.27	2.20	2.11	2.02
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.10	2.03	1.94	1.84
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.93	1.86	1.76	1.66
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.78	1.70	1.59	1.47

جدول XI-القيم الحرجة لاختبار ليليفورز

Table A15 Quantiles of the Lilliefors Test Statistic for Normality[a]

Sample size n	$p = .80$	$.85$	$.90$	$.95$	$.99$
4	.300	.319	.352	.381	.417
5	.285	.299	.315	.337	.405
6	.265	.277	.294	.319	.364
7	.247	.258	.276	.300	.348
8	.233	.244	.261	.285	.331
9	.223	.233	.249	.271	.311
10	.215	.224	.239	.258	.294
11	.206	.217	.230	.249	.284
12	.199	.212	.223	.242	.275
13	.190	.202	.214	.234	.268
14	.183	.194	.207	.227	.261
15	.177	.187	.201	.220	.257
16	.173	.182	.195	.213	.250
17	.169	.177	.189	.206	.245
18	.166	.173	.184	.200	.239
19	.163	.169	.179	.195	.235
20	.160	.166	.174	.190	.231
25	.142	.147	.158	.173	.200
30	.131	.136	.144	.161	.187
Over 30	$\frac{.736}{\sqrt{n}}$	$\frac{.768}{\sqrt{n}}$	$\frac{.805}{\sqrt{n}}$	$\frac{.886}{\sqrt{n}}$	$\frac{1.031}{\sqrt{n}}$

SOURCE. Adapted from Table 1 of Lilliefors (1967), with corrections.

[a] The entries in this table are the approximate quantiles w_p of the Lilliefors test statistic T_1 as defined by Equation 6.2.4. Reject H_0 at the level α if T_1 exceeds $w_{1-\alpha}$ for the particular sample size n.

جدول X- القيم الحرجة لاختبار كولموغوروف- سميرنوف

Quantiles of the Kolmogorov test statistic

One-sided test Two-sided test	$p = 0.90$ $p = 0.80$	0.95 0.90	0.975 0.95	0.99 0.98	0.995 0.99
$n = 1$	.900	.950	.975	.990	.995
2	.684	.776	.842	.900	.929
3	.565	.636	.708	.785	.829
4	.493	.565	.624	.689	.734
5	.447	.509	.563	.627	.669
6	.410	.468	.519	.577	617
7	.381	.436	.483	.538	.576
8	.358	.410	.454	.507	.542
9	.339	.387	.430	.480	.513
10	.323	.369	.409	.457	.489
11	.308	.352	.391	.437	.468
12	.296	.338	.375	.419	.449
13	.285	.325	.361	.404	.432
14	.275	.314	.349	.390	.418
15	.266	.304	.338	.377	.404
16	.258	.295	.327	.366	.392
17	.250	.286	.318	.355	.381
18	.244	.279	.309	.346	.371
19	.237	.271	.301	.337	.361
20	.232	.265	.294	.329	.352
21	.226	.259	.287	.321	.344
22	.221	.253	.281	.314	.337
23	.216	.247	.275	.307	.330
24	.212	.242	.269	.301	.323
25	.208	.238	.264	.295	.317
26	.204	.233	.259	.290	.311
27	.200	.229	.254	.284	.305
28	.197	.225	.250	.279	.300
29	.193	.221	.246	.275	.295
30	.190	.218	.242	.270	.290
31	.187	.214	.238	.266	.285
32	.184	.211	.234	.262	.281
33	.182	.208	.231	.258	.277
34	.179	.205	.227	.254	.273
35	.177	.202	.224	.251	.269
36	.174	.199	.221	.247	.265
37	.172	.196	.218	.244	.262
38	.170	.194	.215	.241	.258
39	.168	.191	.213	.238	.255
40	.165	.189	.210	.235	.252
Approximation for $n > 40$:	$\frac{1.07}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.52}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Source: L. H. Miller. "Table of Percentage Points of Kolmogorov Statistics." *J. Amer. Statist. Assoc.*, 51 (1956), 111–121.

Table K. Table of Critical Values of U in the Mann-Whitney Test* (*Continued*)

Table K$_{III}$. Critical Values of U for a One-tailed Test at $\alpha = .025$ or for a Two-tailed Test at $\alpha = .05$

n_1 \ n_2	9	10	11	12	13	14	15	16	17	18	19	20
1												
2	0	0	0	1	1	1	1	1	2	2	2	2
3	2	3	3	4	4	5	5	6	6	7	7	8
4	4	5	6	7	8	9	10	11	11	12	13	13
5	7	8	9	11	12	13	14	15	17	18	19	20
6	10	11	13	14	16	17	19	21	22	24	25	27
7	12	14	16	18	20	22	24	26	28	30	32	34
8	15	17	19	22	24	26	29	31	34	36	38	41
9	17	20	23	26	28	31	34	37	39	42	45	48
10	20	23	26	29	33	36	39	42	45	48	52	55
11	23	26	30	33	37	40	44	47	51	55	58	62
12	26	29	33	37	41	45	49	53	57	61	65	69
13	28	33	37	41	45	50	54	59	63	67	72	76
14	31	36	40	45	50	55	59	64	67	74	78	83
15	34	39	44	49	54	59	64	70	75	80	85	90
16	37	42	47	53	59	64	70	75	81	86	92	98
17	39	45	51	57	63	67	75	81	87	93	99	105
18	42	48	55	61	67	74	80	86	93	99	106	112
19	45	52	58	65	72	78	85	92	99	106	113	119
20	48	55	62	69	76	83	90	98	105	112	119	127

* Adapted and abridged from Tables 1, 3, 5, and 7 of Auble, D. 1953. Extended tables for the Mann-Whitney statistic. *Bulletin of the Institute of Educational Research at Indiana University*, 1, No. 2, with the kind permission of the author and the publisher.

TABLE K. TABLE OF CRITICAL VALUES OF U IN THE MANN-WHITNEY TEST* (*Continued*)

Table K_{IV}. Critical Values of U for a One-tailed Test at $\alpha = .05$ or for a Two-tailed Test at $\alpha = .10$

n_2 \ n_1	9	10	11	12	13	14	15	16	17	18	19	20
1											0	0
2	1	1	1	2	2	2	3	3	3	4	4	4
3	3	4	5	5	6	7	7	8	9	9	10	11
4	6	7	8	9	10	11	12	14	15	16	17	18
5	9	11	12	13	15	16	18	19	20	22	23	25
6	12	14	16	17	19	21	23	25	26	28	30	32
7	15	17	19	21	24	26	28	30	33	35	37	39
8	18	20	23	26	28	31	33	36	39	41	44	47
9	21	24	27	30	33	36	39	42	45	48	51	54
10	24	27	31	34	37	41	44	48	51	55	58	62
11	27	31	34	38	42	46	50	54	57	61	65	69
12	30	34	38	42	47	51	55	60	64	68	72	77
13	33	37	42	47	51	56	61	65	70	75	80	84
14	36	41	46	51	56	61	66	71	77	82	87	92
15	39	44	50	55	61	66	72	77	83	88	94	100
16	42	48	54	60	65	71	77	83	89	95	101	107
17	45	51	57	64	70	77	83	89	96	102	109	115
18	48	55	61	68	75	82	88	95	102	109	116	123
19	51	58	65	72	80	87	94	101	109	116	123	130
20	54	62	69	77	84	92	100	107	115	123	130	138

* Adapted and abridged from Tables 1, 3, 5, and 7 of Auble, D. 1953. Extended tables for the Mann-Whitney statistic. *Bulletin of the Institute of Educational Research at Indiana University*, **1**, No. 2, with the kind permission of the author and the publisher.

TABLE A-8 Critical Values of T for the Wilcoxon Signed-Ranks Test

	α			
n	.005 (one tail) .01 (two tails)	.01 (one tail) .02 (two tails)	.025 (one tail) .05 (two tails)	.05 (one tail) .10 (two tails)
5	*	*	*	1
6	*	*	1	2
7	*	0	2	4
8	0	2	4	6
9	2	3	6	8
10	3	5	8	11
11	5	7	11	14
12	7	10	14	17
13	10	13	17	21
14	13	16	21	26
15	16	20	25	30
16	19	24	30	36
17	23	28	35	41
18	28	33	40	47
19	32	38	46	54
20	37	43	52	60
21	43	49	59	68
22	49	56	66	75
23	55	62	73	83
24	61	69	81	92
25	68	77	90	101
26	76	85	98	110
27	84	93	107	120
28	92	102	117	130
29	100	111	127	141
30	109	120	137	152

NOTES:

1. * indicates that it is not possible to get a value in the critical region.
2. Reject the null hypothesis if the test statistic T is less than or equal to the critical value found in this table. Fail to reject the null hypothesis if the test statistic T is greater than the critical value found in the table.

TABLE N. TABLE OF PROBABILITIES ASSOCIATED WITH VALUES AS LARGE AS OBSERVED VALUES OF χ_r^2 IN THE FRIEDMAN TWO-WAY ANALYSIS OF VARIANCE BY RANKS*

Table N_{I}. $k = 3$

$N = 2$		$N = 3$		$N = 4$		$N = 5$	
χ_r^2	p	χ_r^2	p	χ_r^2	p	χ_r^2	p
0	1.000	.000	1.000	.0	1.000	.0	1.000
1	.833	.667	.944	.5	.931	.4	.954
3	.500	2.000	.528	1.5	.653	1.2	.691
4	.167	2.667	.361	2.0	.431	1.6	.522
		4.667	.194	3.5	.273	2.8	.367
		6.000	.028	4.5	.125	3.6	.182
				6.0	.069	4.8	.124
				6.5	.042	5.2	.093
				8.0	.0046	6.4	.039
						7.6	.024
						8.4	.0085
						10.0	.00077

$N = 6$		$N = 7$		$N = 8$		$N = 9$	
χ_r^2	p	χ_r^2	p	χ_r^2	p	χ_r^2	p
.00	1.000	.000	1.000	.00	1.000	.000	1.000
.33	.956	.286	.964	.25	.967	.222	.971
1.00	.740	.857	.768	.75	.794	.667	.814
1.33	.570	1.143	.620	1.00	.654	.889	.865
2.33	.430	2.000	.486	1.75	.531	1.556	.569
3.00	.252	2.571	.305	2.25	.355	2.000	.398
4.00	.184	3.429	.237	3.00	.285	2.667	.328
4.33	.142	3.714	.192	3.25	.236	2.889	.278
5.33	.072	4.571	.112	4.00	.149	3.556	.187
6.33	.052	5.429	.085	4.75	.120	4.222	.154
7.00	.029	6.000	.052	5.25	.079	4.667	.107
8.33	.012	7.143	.027	6.25	.047	5.556	.069
9.00	.0081	7.714	.021	6.75	.038	6.000	.057
9.33	.0055	8.000	.016	7.00	.030	6.222	.048
10.33	.0017	8.857	.0084	7.75	.018	6.889	.031
12.00	.00013	10.286	.0036	9.00	.0099	8.000	.019
		10.571	.0027	9.25	.0080	8.222	.016
		11.143	.0012	9.75	.0048	8.667	.010
		12.286	.00032	10.75	.0024	9.556	.0060
		14.000	.000021	12.00	.0011	10.667	.0035
				12.25	.00086	10.889	.0029
				13.00	.00026	11.556	.0013
				14.25	.000061	12.667	.00066
				16.00	.0000036	13.556	.00035
						14.000	.00020
						14.222	.000097
						14.889	.000054
						16.222	.000011
						18.000	.0000006

* Adapted from Friedman, M. 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. Statist. Ass.*, **32**, 688–689, with the kind permission of the author and the publisher.

TABLE N. TABLE OF PROBABILITIES ASSOCIATED WITH VALUES AS LARGE AS OBSERVED VALUES OF χ_r^2 IN THE FRIEDMAN TWO-WAY ANALYSIS OF VARIANCE BY RANKS* (*Continued*)

Table N_{II}. $k = 4$

$N = 2$		$N = 3$		$N = 4$			
χ_r^2	p	χ_r^2	p	χ_r^2	p	χ_r^2	p
.0	1.000	.2	1.000	.0	1.000	5.7	.141
.6	.958	.6	.958	.3	.992	6.0	.105
1.2	.834	1.0	.910	.6	.928	6.3	.094
1.8	.792	1.8	.727	.9	.900	6.6	.077
2.4	.625	2.2	.608	1.2	.800	6.9	.065
3.0	.542	2.6	.524	1.5	.754	7.2	.054
3.6	.458	3.4	.446	1.8	.677	7.5	.052
4.2	.375	3.8	.342	2.1	.649	7.8	.036
4.8	.208	4.2	.300	2.4	.524	8.1	.033
5.4	.167	5.0	.207	2.7	.508	8.4	.019
6.0	.042	5.4	.175	3.0	.432	8.7	.014
		5.8	.148	3.3	.389	9.3	.012
		6.6	.075	3.6	.355	9.6	.0069
		7.0	.054	3.9	.324	9.9	.0062
		7.4	.033	4.5	.242	10.2	.0027
		8.2	.017	4.8	.200	10.8	.0016
		9.0	.0017	5.1	.190	11.1	.00094
				5.4	.158	12.0	.000072

* Adapted from Friedman, M. 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. Statist. Ass.*, **32**, 688–689, with the kind permission of the author and the publisher.

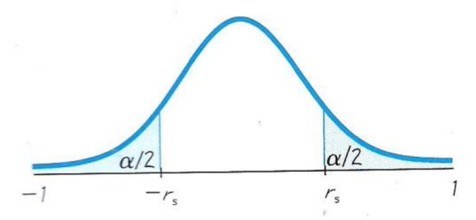


TABLE A-9 Critical Values of Spearman's Rank Correlation Coefficient r_s

n	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.02$	$\alpha = 0.01$
5	.900	—	—	—
6	.829	.886	.943	—
7	.714	.786	.893	.929
8	.643	.738	.833	.881
9	.600	.700	.783	.833
10	.564	.648	.745	.794
11	.536	.618	.709	.755
12	.503	.587	.678	.727
13	.484	.560	.648	.703
14	.464	.538	.626	.679
15	.446	.521	.604	.654
16	.429	.503	.582	.635
17	.414	.485	.566	.615
18	.401	.472	.550	.600
19	.391	.460	.535	.584
20	.380	.447	.520	.570
21	.370	.435	.508	.556
22	.361	.425	.496	.544
23	.353	.415	.486	.532
24	.344	.406	.476	.521
25	.337	.398	.466	.511
26	.331	.390	.457	.501
27	.324	.382	.448	.491
28	.317	.375	.440	.483
29	.312	.368	.433	.475
30	.306	.362	.425	.467

NOTES: For $n > 30$, use $r_s = \pm z/\sqrt{n-1}$, where z corresponds to the level of significance.

For example, if $\alpha = 0.05$, then $z = 1.96$.

If the absolute value of the test statistic r_s exceeds the positive critical value, then reject H_0: $\rho_s = 0$ and conclude that there is a correlation.

Based on data from "*Biostatistical Analysis,* 4th edition," © 1999, by Jerrold Zar, Prentice Hall, Inc., Upper Saddle River, New Jersey.

TABLE A-6 Critical Values of the Pearson Correlation Coefficient r

n	$\alpha = .05$	$\alpha = .01$
4	.950	.999
5	.878	.959
6	.811	.917
7	.754	.875
8	.707	.834
9	.666	.798
10	.632	.765
11	.602	.735
12	.576	.708
13	.553	.684
14	.532	.661
15	.514	.641
16	.497	.623
17	.482	.606
18	.468	.590
19	.456	.575
20	.444	.561
25	.396	.505
30	.361	.463
35	.335	.430
40	.312	.402
45	.294	.378
50	.279	.361
60	.254	.330
70	.236	.305
80	.220	.286
90	.207	.269
100	.196	.256

NOTE: To test H_0: $\rho = 0$ against H_1: $\rho \neq 0$, reject H_0 if the absolute value of r is greater than the critical value in the table.

قيم إحتمال الدلالة p المقابلة لقيم Z المعيارية

Table A. Table of Probabilities Associated with Values as Extreme as Observed Values of z in the Normal Distribution

The body of the table gives one-tailed probabilities under H_0 of z. The left-hand marginal column gives various values of z to one decimal place. The top row gives various values to the second decimal place. Thus, for example, the one-tailed p of $z \geq .11$ or $z \leq -.11$ is $p = .4562$.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
3.2	.0007									
3.3	.0005									
3.4	.0003									
3.5	.00023									
3.6	.00016									
3.7	.00011									
3.8	.00007									
3.9	.00005									
4.0	.00003									

هنا شكل جديد لجدول Z المعياري مع الاحتمال P

Z

TABLE E. TABLE OF CRITICAL VALUES OF D IN THE KOLMOGOROV-SMIRNOV ONE-SAMPLE TEST*

Sample size (N)	Level of significance for D = maximum $\lvert F_0(X) - S_N(X)\rvert$				
	.20	.15	.10	.05	.01
1	.900	.925	.950	.975	.995
2	.684	.726	.776	.842	.929
3	.565	.597	.642	.708	.828
4	.494	.525	.564	.624	.733
5	.446	.474	.510	.565	.669
6	.410	.436	.470	.521	.618
7	.381	.405	.438	.486	.577
8	.358	.381	.411	.457	.543
9	.339	.360	.388	.432	.514
10	.322	.342	.368	.410	.490
11	.307	.326	.352	.391	.468
12	.295	.313	.338	.375	.450
13	.284	.302	.325	.361	.433
14	.274	.292	.314	.349	.418
15	.266	.283	.304	.338	.404
16	.258	.274	.295	.328	.392
17	.250	.266	.286	.318	.381
18	.244	.259	.278	.309	.371
19	.237	.252	.272	.301	.363
20	.231	.246	.264	.294	.356
25	.21	.22	.24	.27	.32
30	.19	.20	.22	.24	.29
35	.18	.19	.21	.23	.27
Over 35	$\frac{1.07}{\sqrt{N}}$	$\frac{1.14}{\sqrt{N}}$	$\frac{1.22}{\sqrt{N}}$	$\frac{1.36}{\sqrt{N}}$	$\frac{1.63}{\sqrt{N}}$

* Adapted from Massey, F. J., Jr. 1951. The Kolmogorov-Smirnov test for goodness of fit. *J. Amer. Statist. Ass.*, **46**, 70, with the kind permission of the author and publisher.